

STIC-Biotech/ChemLib

88508

From: Yu, Misook
Sent: Saturday, March 08, 2003 1:08 PM
To: STIC-Biotech/ChemLib
Subject: 09/502,945

Please search proteins encoded by SEQ ID NOs 1-5. Thank you,

Examiner Misook Yu, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)

RECEIVED
MAR 10 2003
STIC

Edward Han
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 3/11/03
Date Completed: 3/25/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH: 5 - reverse to FA
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 03
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run On: March 21, 2003, 12:20:42 ; Search time 98.9379 Seconds
(without alignments)
12016.549 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346

Sequence: 1 ggaattcctctgtcgaagt.....aaacaaaagttaaaattt 2885

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p.model -DEV=xl
-Q-/cgn2_1/USPTO.spool/US09502945/runat_14032003_101058_19113/app_query.fasta_1.10979
-DB-SPTREMBL_21 -OFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09502945@cgn_1_1565_erunat_14032003_101058_19113 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	1831.5	34.3	878	4	Q9UFU7	Q9ufu7 homo sapien

ID	Q9UFU7	PRELIMINARY;	PRT;	878 AA.
AC	Q9UFU7			
DT	01-MAY-2000 (TREMBlrel. 13, Created)			
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)			
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)			
DE	Hypothetical 94.3 kDa protein (Fragment).			
GN	DKFZP586J0917			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=UTERUS;			
RA	Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;			
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL117455; CAB55935.1; ..			
DR	InterPro: IPR000286; His_deacetylase.			
DR	Pfam: PF00850; Hist_deacetyl; 1.			
DR	PRINTS; PR01270; HDASUPER.			
KW	Hypothetical protein.			
FT	NON_TER 1			
SQ	SEQUENCE 878 AA; 94330 MW; 26882E693960E10D CRC64;			

ALIGNMENTS

RESULT 1

Q9UFU7
ID Q9UFU7
AC Q9UFU7
DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical 94.3 kDa protein (Fragment).

GN DKFZP586J0917

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=UTERUS;

RA Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL117455; CAB55935.1; ..

DR InterPro: IPR000286; His_deacetylase.

DR Pfam: PF00850; Hist_deacetyl; 1.

DR PRINTS; PR01270; HDASUPER.

KW Hypothetical protein.

FT NON_TER 1

SQ SEQUENCE 878 AA; 94330 MW; 26882E693960E10D CRC64;


```

QY 1907 TTTGCCATCATCGGCCCCCAGGACACCCAGCGAGGAATCCACAGCCATGGGATCTGC 1966
Db 1908 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
587 PheAlaValAlaArgProProGlyHisHisAlaAspHisSerThrAlaMetGlyPheCys 606
QY 1967 TTTCTCAACTCTGTAGCATTACCCAGCAAACTCTCTACAGCAGAGTGAACGTGGCAAG 2026
Db 1968 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
607 PhePheAsnSerValAlaIleAlaCysArgGlnLeuGlnGlnGlnSerLysAlaSerLys 626
QY 2027 GTCTCTCATCTGGACTGGGACATTCACATGGCAATGGCACCCAGCAGGCGTTCTACAAT 2086
Db 2028 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
627 IleLeuIleValAlaAspTrpAspValHisHisGlyAsnGlyThrGlnGlnThrPheTyrGln 646
QY 2087 GACCCCTCTGTCTACATCTCTCTCCATCGCTATGACACGGGAACTTCTTCCAGGC 2146
Db 2088 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
647 AspProSerValLeuTyrIleSerLeuHisArgHisAspAspGlyAsnPhePheProGly 666
QY 2147 TCTGGGCTCTCTGAAGAGTTGGTGGAGCAGGAGGCGTGGTACATGTGAACGTGGCA 2206
Db 2148 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
667 SerGlyAlaValAlaAspGluValGlyAlaGlySerGlyGluGlyPheAsnValAlaAla 686
QY 2207 TGGACAGGAGTGTGGACCCGCCATTTGGAGAGCTGGAGTACCTTACAGCCTTCAGGACA 2266
Db 2208 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
687 TrpAlaGlyGlyLeuAspProPrometGlyAspProGluTyrLeuAlaAlaPheArgIle 706
QY 2367 GTGCTGATGCCATTGCCAGAGTTCTACCTGATGTGTCTAGTCTCCGCGGGTTT 2326
Db 2368 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 ValValMetProIleAlaArgGluPheSerProAspLeuValLeuValSerAlaGlyPhe 726
QY 2327 GATGCTGTTGAAGACATCTGTCTCTGGTGGTCTCTACCTGAGGCGGCGGTGTGCTGG 2386
Db 2328 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 AspAlaAlaGluGlyHisProAlaProLeuGlyGlyTyrHisValSerAlaLysCysPhe 746
QY 2387 GGCCTTGTACAGGAGCTGATGACCTGGCAGGCGGCGGTGTGCTGGCGCTGGAG 2446
Db 2388 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
747 GlyTyrMetThrGlnGlnLeuMetAsnLeuAlaGlyGlyAlaValValLeuAlaLeuGlu 766
QY 2447 GGAGGCGATGACATTGACCGCATCTGTGATGCTCTGAGCTGTGCTCGGCTCTGCTC 2506
Db 2448 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
767 GlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValAlaAlaLeuLeu 786
QY 2507 AGTGTAAAGCTGAGCCCTTGGATGAGCAGTCTTCCAGCAAAAGCCCAACATCAACGCA 2566
Db 2508 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
787 GlyAsnArgValAlaAspProLeuSerGluGluGlyTrpLysGlnLysProAsnLeuAsnAla 806
QY 2567 GTGGCCACGCTAGAGAAATCATCGATCCAGATCCAGAGCAACACTGGAGCTGTGTCAGAA 2626
Db 2568 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
807 IleArgSerLeuGluAlaValIleArgValHisSerLysTyrTrpGlyCysMetGlnArg 826
QY 2627 TTCGCGCTGGTCTGGCGGTCCTCGCGAGGCGGCCACAGCAGGTGAGACCGAAGCC 2686
Db 2628 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
827 LeuAlaSerCysProAspSerTrpValProArgValProGlyAlaAspLysGluGluVal 846
QY 2687 GAAAT-GTGAACGCCATGGCTTGTCTGTGTGGTGGG-----GCCGAACAGGCCCAAGCT 2739
Db 2688 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
847 GluAlaValThrAlaLeuAlaSerLeuSerValGlyIleLeuAlaGluAsp----- 863
QY 2740 CGGGACCGCGGGAACACAGCCCCAGCGCGGAGAGGCCATGGAGCAGGAG 2793
Db 2741 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
864 -----ArgProSerGluGlnLeuValGluGluGlu 873

```

RESULT 2
Q9JL72

ID Q9JL72 PRELIMINARY; PRT; 938 AA.
AC
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Histone deacetylase 7.
GN HDAC7A OR HDAC7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6.C57BL/6;
RX MEDLINE=20107033; PubMed=10640276;
RA Kao H.Y., Downes M., Ordentlich P., Evans R.M.;
RT "Isolation of a novel histone deacetylase reveals that class I and
RL class II deacetylases promote SMRT-mediated repression.";
RL Genes Dev. 14:55-66(2000).
DR EMBL: AF207749; AAF31419.1; -
DR MGD; MGI:1891835; Hdac7a.
DR InterPro; IPR000286; Hist_deacetylase.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PRO1270; HDASUPER.
SQ SEQUENCE 938 AA; 101318 MW; A282A36BEC840897 CRC64;

Alignment Scores:
Pred. No.: 8,69e-99 Length: 938
Score: 1773.00 Matches: 422
Percent Similarity: 54.79% Conservative: 110
Best Local Similarity: 43.46% Mismatches: 250
Query Match: 33.16% Indels: 190
DB: 11 Gaps: 24

US-09-502-945-2 (1-2885) x Q9JL72 (1-938)
QY 2 GAATTCCTCTTGTGGAAGTCAAGGAGGAGCCACACAGGCGGCTCAACCATTCCTCCCA 61
Db 114 GluValIleLeuLysGlnGlnAlaLeuGluArgThrValHisProSerSerPro.133
QY 62 CAGCACCC- - - - - AAATGCTGGGAGCCCAACCATCTCTTTTG. 100
Db 134 SerIleProTyrArgThrLeuGluProLeuAspThrGluGlyAlaAlaArgSerValLeu.153
QY 101 GACCAGAGTTCCTCCCGCAGAGCGGCCCTCGGAGCCCTCCCTCTCTACAACTGCCT 160
Db 154 SerSerPheLeuProValProSerLeuProThrGluProProGluHis - - - - - 170
QY 161 TTGCTGGCCCTACGACAGTCCGAGCAGTCTCCCTCCGCAAAACAGCCTCTGAACCC 220
Db 171 - - - - - PheProLeuArgLysThrValSerGluPro.180
QY 221 AACTTGAAGTCGTTCAAGGCTAAACAGAGGTGGCTGACGAGAGAGAGTCCCTTC 280
Db 181 AsnLeuLysLeuArgTyrLysProLys - - - - - LysSerLeuGluArgGlyAsnProLeu.199
QY 281 CTGCGCTGCGAAGGATGGGACTCTTATTAGCAGCTTTAAGAGAGAGCTGTTGAGATCACA 340
Db 200 LeuArgLysGluSerAlaPro - - - - - ProSerLeuArgArgProAlaGluThrLeu.217
QY 341 GGTGCGGCGCTGGGCGCTGCTGCTGTGTATACAGCGCAGCCGCTCGGCGCCAGCTCT 400
Db 218 GlyAspSerSerProSerSerSer - - - - - SerThrProAlaSerGlyCysSerSer.234
QY 401 CCCACAGCTCCACAGCAGCACCATCCCTGAGATGCTTACTGCTCAGTCCCAACATC 460
Db 235 ProAsnAspSer - - - - - GluHisGly - - - - - 241
QY 461 CCCACTGAGATCTCTCTCAGCAGCGGCCCTCTCTGGACAGTCCCTCCCAACAGCTTC 520
Db 241 - - - - - 241
QY 521 AGCCTCTACAGTCTCTCTTCTGTGCCCAACATCTCCCTAGGCTGCAGGCCAGGCTCACT 580
Db 242 - - - - - ProAsnProAlaLeuGlySerGluAlaAspGlyAsp.253
QY 581 GTCACCAACTCAGCAGTCACTGCTCCCTCCCGAAG - - - - - 613
Db 254 - - - - - ArgArgThrHisSerThrLeuGlyProArgGlyProValLeuGlyAsnProHisAla.272
QY 614 - - - - - CTGTGCGACACAGCAGGAGCGCCGAGGAGCGCCCTCCAGTCCCTCGGCGCAGGTTGGC 670
Db 273 ProLeuPheLeuHisHisGlyLeuGluProGluAla - - - - - GlyGly 286

```


QY	671	ACGCTGACGGCAAGTTCATGAGC	ATCCTCTATTCTGGCTGCCTGCGCGTGGCA	730
Db	287	ThrLeu		288
QY	731	CTGAGGGCGAGCGGAGCC	CCACGGCGATGCCTCCCTGCTGCGAGCATGCTGTGCTG	790
Db	289		ProSerArgLeuGlnProIleLeuLeuLeu	298
QY	791	GAGCAGGCGCGCAGCAGAGC	ACCTCCTCTGCTGG	832
Db	299	AspProSerValSerHisAla	ProLeuThrValProGlyLeuGlyProLeuProPhe	318
QY	833	CAGGGCAGTCCCACTAGTACG	GGGTGAAGCTGTGGCCACCAAGCATGCGGACGGTAGGC	892
Db	319	HisPheAlaGlnProLeuLeuThr	ThrGluArgLeuSerGlySer	333
QY	893	AAGCTCCGGGCATCGCCCTG	AGCGCAGCTACGTCTCACCGCTGCCGACAGTCCC	952
Db	334		GlyLeuHisArgProLeuAsnArgThrArgSerGluProLeuProProSerAla	351
QY	953	CAGGCCCTGCAGCAGCTGTG	TCATGCAACACAGCACCAGCAGTAGTCTCCCTGAGAGCAGAAG	1012
Db	352	ThrAlaSerProLeuLeuAla	ProLeuGlnProArgGlnAspArgLeu	367
QY	1013	CAGCAGCAGCTACAGCTGG	CGCAAGATCCTCACCAGACAGGGGAGTGC	1072
Db	368	LysProHisValGlnLeuIleLys	ProAlaIleSerProProGlnArgProAlaLysPro	387
QY	1073	ACCACCCACCCCTGAGGAG	CACAGAGGAGTGCAGCGAGCAGCAGGAGGTCTTCTGTGGG	1132
Db	388	SerGlnLysProArgLeuArg	GlnIleProSerAlaGluAspLeuGluThr	407
QY	1133	GAGGAGCCCTGACCATGCTG	CCCCGGGAGGCTC--CACAGAGAGTGAGACACAGAGAA	1189
Db	407	LyGlyValGlyProMetAlaAsn	AspGlyLeuGluHisArgGlu	424
QY	1190	GACCTGGAGGAGGAGCAGC	AGGAGNAGGNTGGGAGGAGGAGGAGTGCATCCAGTT	1249
Db	424	rgGlyProProGluGlyArgGly	SerIleSerLeuGlnGlnHisGlnValProPro	443
QY	1250	AAGCAGGAGGAGGCGCAG	TGTGTGTGAGGAGGGCCGCTACTTGTAGGAGCTGTGTGT	1309
Db	444	TrpGlnGlnHisLeuAlaGly	Arg	458
QY	1310	GGATACAAAACATGTCTC	AGATGCCCAACCGCTGCACCTTGCAGGTGTACCAAGC	1369
Db	459	GlyAspSerValLeu		463
QY	1370	CCCTCAGCTGCGCCACTGT	CCCCACCAAGCCCTGGCGCTACCAATCTCCCTGTCT	1429
Db	464		IleProLeuAlaGlnValGlyHisArgProLeuSerArgThrGlnSerSerProAla	482
QY	1430	GCCCTGGGGGCATGAAGA	ACCCCCCAGACCCCGTC--	1468
Db	483	AlaProValSerMetLeuSer	Pro	500
QY	1469		AAGCACCCTCTTACCACAAGTGTGGTCTACGACACGTTCACTGAACACCAAGTGC	1525
Db	501	SerSerGluThrProAlaThr	GlyLeuValTyraSpSerValMetLeuLysHisGlnCys	520
QY	1526	ATGTGCGGGAACACACACG	TGCACCTGAGCATGTGCCGGATCCAGAGCATCTGTGTCC	1585
Db	521	SerCysGlyAspAsnSerLys	HisProGluHisAlaGlyArgIleGlnSerIleArgSer	540
QY	1586	CGGCTGAGGAGCAGCGCTG	CTTACCAAGTCGCGAGCGGATCCGAGGTGCGAAAGCCACG	1645
Db	541	ArgLeuGlnGluArgGlyLeu	ArgSerGlnCysGluCysLeuArgGlyArgLysAlaSer	560
QY	1646	CTAGATGAGATCAGACAGT	GCACTCTGATACCACACCCCTGCTCTATGGGACCAAGTCCC	1705
Db	561	LeuGluGluLeuGlnSerVal	HisSerGluArgHisValLeuLeuTyrglyThrAsnPro	580

Qy	1706	CTCAACGGCGAAGCTAGACAGCAAGAAGTTGCTCGTCCCATCCGACAGAGATGTAT	17165
Db	581	LeuSerArgLeuLysLeuAspAsnGlyLysLeuThrGlyLeuLeuAlaGlnArgThrPhe	600
Qy	1766	GCTGTGCTGCTTGCGGGGCATCGGGGTGGACAGTACACCGTGTGGAATAGATGCAC	1825
Db	601	ValMetLeuProCysGlyValGlyValAspThrAspThrIleIleTrpAsnGluLeuHis	620
Qy	1826	TCCTCCAGTCTTCGGCGCATCGGCGTGGCTGCCTCTGGAGCTGGCCCTCAAGGTGGCT	1885
Db	621	SerSerAsnAlaAlaArgTrpAlaAlaGlySerValThrAspLeuAlaPheLysValAla	640
Qy	1886	GCAGGAGAGCTCAAGAAATGGATTGCCATCATCCGGCCCCCAGACACCACCGCCGAGAA	1945
Db	641	SerArgGluLeuLysAsnGlyPheAlaValValArgProGlyHisHisAlaAspHis	660
Qy	1946	TCCACACCATGGATTCTGCTCTTCAACTCTGTAGCCATCACCGCAAAACTCTCTACAG	2005
Db	661	SerThrAlaMetGlyPheCysPheAsnSerValAlaIleAlaCysArgGlnLeuGln	680
Qy	2006	CAGAAGTTGAACCTGGCGCAAGTCTCATCTGCTGGACTGGGACATTCACCATGGCAATGGC	2065
Db	681	GlnHisGlyLysAlaSerLysIleLeuIleValAspTrpAspValHisHisGlyAsnGly	700
Qy	2066	ACCAGCAGCGGTTCTACAAATGACCCTCTGTGCTCTACATCTCTCTGCATCGCTATGAC	2125
Db	701	ThrGlnGlnThrPheTyrGlnAspProSerValLeuTyrIleSerLeuHisArgHisAsp	720
Qy	2126	AACGGGAACCTCTTCCAGGCTCTGGGCTCCTGAAGAGTTGTTGGAGGACCAAGCGTGT	2185
Db	721	AspGlyAsnPhePheProGlySerGlyAlaValAspGluValGlyThrAlaSerGlyGlu	740
Qy	2186	GGGTACAATGTGAACGTGGCATCGACAGCAGAGGTGTGGACCCCCCATCTGGAGAGCTGG	2245
Db	741	GlyPheAsnValAsnValAlaTrpAlaGlyGlyLeuAspProPrometGlyAspProGlu	760
Qy	2246	TACCTTACAGCCTTACAGCAGTGTGTATGCCATTTGCCAGTGTATGCCAGGAGTCTCACT	2305
Db	761	TyrLeuAlaAlaPheArgIleValValMetProIleAlaArgGluPheAlaProAspLeu	780
Qy	2306	CTCTAGTCTCCGCGGTTTGATGCTGTGAAGGACATCTGCTCPCCTGGGTGGCTAC	2365
Db	781	ValLeuValSerAlaGlyPheAspAlaAlaGluGlyHisProAlaProLeuGlyGlyTyr	800
Qy	2366	TCTGTCAACCGCAGATGTTTTGGCCACTTTGACCAGCAGCTGATGACCTCGCAGGGGCG	2425
Db	801	HisValSerAlaLysCysPheGlyTyrMetThrGlnGlnLeuMetAsnLeuAlaGlyGly	820
Qy	2426	CGGTGTGTGTCGCCCTGGAGGAGGCCATGACTTGACCGCCATCTGTGATGCCCTCTGAA	2485
Db	821	AlaValValLeuAlaLeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGlu	840
Qy	2486	GCTTGTGTCTCGGCTCTGCTCAGTGTAAAGTGTACAGCCCTTGGATGAGGCAGTCTTGAC	2545
Db	841	AlaCysValAlaAlaLeuLeuGlyAsnLysValAspProLeuSerGluGluSerTrpLys	860
Qy	2546	CAAAAGCCCAACATCAACGCGATGGCCAGCTAGAGAAAGTCATCCAGATCCAGAGCAAA	2605
Db	861	GlnLysProAsnLeuSerAlaIleArgSerLeuGluAlaValValArgValHisArgLys	880
Qy	2606	CACGTGAGCTGTGTGCAGAGTTTCGCCGC-----TGTCTGGCGCGTCCCT	2652
Db	881	TyrTrpGlyCysMetGlnArgLeuAlaSerCysProAspSerTrpLeuProArgValPro	900
Qy	2653	CGCAGGGCCCCACAGGCTGAGACCGGAGGCCGAATGTCAAGCCCATGCCCTTGCTGT	2712
Db	901	GlyAla-----AspAlaGluValGluAlaValThrAlaLeuAlaSerLeu	915
Qy	2713	TTGTGTGGG-----GCCGAACAGGCCCAAGCTGCGGCACCGCGGGAACACAGCCCCAGG	2766
Db	916	SerValGlyIleLeuAlaGluAsp-----Arg	924
Qy	2767	CGCGCAGAGGAGCCCATGTGAGCAGGAG	2793

QY	1487	AGTGTGCTTACGACACAGTTTCATGCTAAAGCACACAGTCAGTCGTCCGGGAACACACACAGCTG	1546
		#####	
DB	481	GlyLeuIleTyrAspSerValMetLeuIysHisGlnCysSerCysGlyAspAsnSerArg	500
		#####	
QY	1547	CACCTTGAGCATGCTGGCGCGATCCAGACATCTGGTCCCGCTGTCAGGAGACAGCGCTG	1606
DB	501	HisProGluHisAlaGlyArgIleGlnSerIleTyrSerArgLeuGlnGluArgGlyLeu	520
QY	1607	CTTAGCAAGTGCAGCGCGATCCAGATGCCAAAGCCAGCTAGATGAGATCCACAGCAGTG	1666
		#####	
DB	521	ArgSerGlnCysGluCysLeuArgGlyArgLysAlaSerLeuGluGlnSerVal	540
QY	1667	CATCTGTAATPACACACCTGCTCTATGGGACCATCTCCCTCAACCGCAGAGACTAGAC	1726
		#####	
DB	541	HisSerGluArgHisValLeuLeuTyrGlyThrAsnProLeuSerArgLeuLysLeuAsp	560
QY	1727	AGCAAGAAGTGTCTGCTCCATCAGCCAGAGAAGATGATGCTGTCTGCTCTGTGGGGGC	1786
		#####	
DB	561	AsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPheGluMetLeuProCysGlyGly	580
QY	1787	ATCGGGGTGACAGTGCACCGCTGTGGAATGAGATGCACACTCTCCAGTGTCTGCCGATG	1846
		#####	
DB	581	ValGlyValAspThrAspThrIleTyrAsnGluLeuHisSerSerAsnAlaAlaArgTyr	600
QY	1847	GCAGTGGGTCCTGCTGGAGCTGGCCCTTCAAGTGTGGTGCAGGAGAGCTCAAGAATGGA	1906
		#####	
DB	601	AlaAlaGlySerValThrAspLeuAlaPheLysValAlaSerArgGluLeuLysAsnGly	620
QY	1907	TTTSCCATCATCCGGCCCCCAGACACACCGCGAGGAATCCACAGCCATCGGATTCCTGC	1966
		#####	
DB	621	PheAlaValValArgProGlyHisHisAlaAspHisSerThrAlaMetGlyPheCys	640
QY	1967	TTCTTCAACTCTGTAGCCATCACCGCAAACTCTACAGCAGAAG-----TTGAAC	2017
		#####	
DB	641	PhePheAsnSerValAlaIleAlaCysArgGlnLeuGlnGlnSerLysAlaSerLys	660
QY	2018	GTGGGCAGGTCTCATCTGTGGACTGGGACATTCACCATGGCAATGGCACCCAGCAGCG	2077
		#####	
DB	661	AlaSerLysIleLeuIleValAspTyrAspAspValHisHisGlyAsnGlyThrGlnGlnThr	680
QY	2078	TTTCACAAATGACCCCTCTGTGCTCATCTCTCTGCATCGCTATGACAAACGGGAACCTC	2137
		#####	
DB	681	PheTyrGlnAspProSerValLeuTyrIleSerLeuHisArgHisAspAspGlyAsnPhe	700
QY	2138	TTTCCAGCTCTGGGCTCTCTGAAGAGTTGTGGAGGACCGGTGGGTGTACAAATGTG	2197
		#####	
DB	701	PheProGlySerGlyAlaValAspGluValGlyAlaGlySerGlyGluGlyPheAsnVal	720
QY	2198	AACGTGGCATGGACAGAGTGTGGACCCCCCCCCCATTTGGAGAGCTGGAGTACCTTACGCC	2257
		#####	
DB	721	AsnValAlaTyrAlaGlyGlyLeuAspProPrometGlyAspProGluTyrLeuAlaAla	740
QY	2258	TTACAGGACAGTGGTGTATGCCCATTTGCCACAGCTTCTCACTGTGTGTCTAGTCTCC	2317
		#####	
DB	741	PheArgIleValValMetProIleAlaArgLysPheSerProAspLeuValLeuValSer	760
QY	2318	CCCGGGTTTGATGCTGTGAAGGACATCTGTCTCTCTGGTGGGTACTCTGTCTACCGCC	2377
		#####	
DB	761	AlaGlyPheAspAlaAlaGluGlyHisProAlaProLeuGlyGlyTyrHisValSerAla	780
QY	2378	AGATGTTTGGCCACTTGACCGACGCTGATCACCTCGCAGGGGGCGGTGTGTGCTG	2437
		#####	
DB	781	LysCysPheGlyTyrMetThrGlnGlnLeuMetCAsnLeuAlaGlyAlaValValLeu	800
QY	2438	GCCTGGAGGAGGCCATGACTTGACCGCCATCTGTGTGATCCCTCTGAAGCTTGTCTCG	2497
		#####	
DB	801	AlaLeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValAla	820
QY	2498	GCTCTGCTCAGTGTAAAGCTGCAGCCCTTGGATGAGGCACTTTCGACGAAAGCCC	2554
		#####	
DB	821	AlaLeuLeuGlyAsnArgValAspProLeuSerGluGluGlyTyrLysGlnLysPro	839

RESULT 4			
ID	Q8WU14	PRELIMINARY;	PRT; 614 AA.
AC	Q8WU14; Q9BR73;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Histone deacetylase 7A.		
OS	homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=COLON;		
RA	Strausberg R.;		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBDJ databases.		
DR	EMBL; BC020505; AAH0505.1; -.		
DR	EMBL; BC006453; AAH06453.1; -.		
DR	InterPro; IPR000286; His_deacetylase.		
DR	Pfam; PF00850; Hist_deacetyl; 1.		
DR	PRINTS; PR01270; HDASUPER.		
SQ	SEQUENCE 614 AA; 66186 MW; 2B638DD8C866B502 CRC64;		
Alignment Scores:			
	Pred. No.:	7.41e-89	Length:
	Score:	1607.50	Matches:
	Percent Similarity:	61.32%	Conservative:
	Best Local Similarity:	49.14%	Mismatches:
	Query Match:	30.07%	Indels:
	DB:	4	Gaps:
US-09-502-945-2 (1-2885) x Q8WU14 (1-614)			
QY	851	GTGACGGGTGAACCTGTGGCCACGACATCGCGAGGTAGGCAAGCTCCCGCGCATCGG	910
		:	
Db	1	MethThrGluArgLeuSerGlySer-----	-----GlyLeuHisTrp 13
QY	911	CCCTGTAGCCGCACCTAGTCTCTCACCGTTCGCGCAGAGTCCCGAGGCC-----	-----958
		: :	
Db	14	ProLeuSerArgThrArgSerGluProLeuProProSerAlaThrAlaProProPro	33
QY	959	-----CTGCAGCAGCTGGTTCATGCACACACACACAGCAGGATTCCTG	1000
		: :	
Db	34	GlyProMetGlnProArgLeuGluGlnLeu-----	-----LysThrHisValGlnValIle 50
QY	1001	GAGAGCAGAGACGACGACGACGTACAGCTGGCGCAGATCCTCCACGACAGAGGGAGCTG	1050
		: :	
Db	51	LysArgSer-----	-----AlaLysProSerGluLys 59
QY	1061	CCCAGCAGCCACCACCCACCTCTGAGGACAGACAGAGGAGCTGACGGACGACGAGGAG	1120
Db	60	ProArg-----	-----61
QY	1121	GTCTTGTGGGGAGGAGGCCCTGACCATGCCCGGGAGGGCTCCACAGAGTGTAGAGC	1180
Db	61	-----	-----61
QY	1181	ACACAGGAAGACCTGGAGGAGGAGGACGAGATGGGAGGAGGAGGAGGATTGC	1240
Db	62	LeuArgGlnIleProSerAlaGluAspLeuGluThrAspGly-----	-----75
QY	1241	ATCCAGGTTAAGGACGAGGAGGGCGAGAGTGGT-----GCTGAGGAGGGGCCGACATTG	1294
Db	76	-----GlyGlyProGlyGlnValValAspGlyLeuGluHis	88
QY	1295	GAGGAGCCTGGTGGTATACAAAAAACTGTTCTCAGATGCCACCGCTGCACACCTTTG	1354

```

Db 89 ArgGluLeuGlyHisGlyGln-----ProGluAlaArgGlyProAlaProLeu 104
QY 1355 CAGGTGTACCAAGCCGCC----- 1372
Db 105 Gln-----GlnHisProGlnValLeuLeuTrpGluGlnArgLeuAlaGlyArgLeu 122
QY 1373 -----CTCAGCCCTGGCCACTGTGTCGCCCAACCAAGCC 1402
Db 123 ProArgGlySerThrGlyAspThrValLeuLeuProLeuAlaGlnGlyHisArgPro 142
QY 1403 CTGGCCCTACCAATCTCCCTGCTGCCCTGGGGCATGAAGAACC----- 1453
Db 143 LeuSerArgAlaGlnSerSerProAlaAlaProAlaSerLeuSerAlaProGluProAla 162
QY 1454 -----CCACACCAACCCGTCAGCACCTC-----TTCCACACA 1486
Db 163 SerGlnAlaArgValLeuSerSerSerGluThrProAlaArgThrLeuProPheThrThr 182
QY 1487 AGTGTGTCTACGACACGTTTCATGCTAAAGCACCATGTCATGTCGGGAACACACACGTCG 1546
Db 183 GlyLeuIleTyrAspSerValMetLeuLysHisGlnCysSerCysGlyAspAsnSerArg 202
QY 1547 CACCTGAGCATGTGGCCGATCCAGAGCATCTGTGTCGCCGCTGCAGGACAGCGCTG 1606
Db 203 HisProGluHisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnArgGlyLeu 222
QY 1607 CTTAGCAAGTCGAGCGGATCCGAGTCCGAAGCCAGCTAGATGATCCAGACAGTGT 1666
Db 223 ArgSerGlnCysGluCysLeuArgGlyArgLysAlaSerLeuGluLeuGlnSerVal 242
QY 1667 CACTCTGAATACCACACCTCTCTATGGGACCATGTCCTCAACCGCAGCAAGCTAGAC 1726
Db 243 HisSerGluArgHisValLeuLeuTyrGlyThrAsnProLeuSerArgLeuLysLeuAsp 262
QY 1727 AGCAAGAGTGTCTGCTCCATCCAGCAGAGATGTATGCTGCTGCTCTGTGGGGC 1786
Db 263 AsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPheValMetLeuProCysGlyGly 282
QY 1787 ATCCGGGTGGACATGACACCGTGTGAATGATGATGATGATGATGATGATGATGATG 1846
Db 283 ValGlyValAspThrAspThrIleTrpAsnGluLeuHisSerAsnAlaAlaArgTrp 302
QY 1847 GCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1906
Db 303 AlaAlaGlySerValThrAspLeuAlaPheLysValAlaSerArgGluLeuLysAsnGly 322
QY 1907 TTGCGCATCATCGGCCGCCAGGACACCCAGCAGCAGGAAATCCAGCATGGATTCGTC 1966
Db 323 PheAlaValValArgProProGlyHisHisAlaAspHisSerThrAlaMetGlyPheCys 342
QY 1967 TTCTTCACTCTGTAGCCATCACCAGCAAACTCTACAGCAGAAGTTGAACGTGGGCAAG 2026
Db 343 PhePheAsnSerValAlaIleAlaCysArgGlnLeuGlnSerLysAlaSerLys 362
QY 2027 GTCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2086
Db 363 IleLeuIleValAspTrpAspValHisHisGlyAsnGlyThrGlnGlnThrPheTyrGln 382
QY 2087 GACCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2146
Db 383 AspProSerValLeuTyrIleSerLeuHisHisAspGlyAsnPhePheProGly 402
QY 2147 TCTGGGCTCTCAAGAGTGTGGGAGGACAGCGGTGGGTGCTGCTGCTGCTGCTGCTGCT 2206
Db 403 SerGlyAlaValAspGluValGlyAlaGlySerGlyGlyGlyPheAsnValAsnValAla 422
QY 2207 TGGACAGGAGGTGTGGACCCCTTACCTGAGAGCTGGAGCTGACCTTACAGCTTCAGACA 2266
Db 423 TrpAlaGlyCysLeuAspProMetGlyAspProGluTyrLeuAlaAlaPheArgIle 442
QY 2267 GTGGTATGCCCATTCGCCACAGTGTCTCACCTGATGTGTCTGCTGCTGCTGCTGCTGCT 2326
Db 443 ValValMetProIleAlaArgGluPheSerProAspLeuValLeuValSerAlaGlyPhe 462

```

```

QY 2327 GATCCTCTTTGAAGACATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2386
Db 463 AspAlaAlaGluGlyHisProAlaProLeuGlyGlyTyrHisValSerAlaLysCysPhe 482
QY 2387 GGCACCTTACCAGGACCTGATGACCTGGCAGGGGGGGGGTGGTGGTGGTGGTGGTGGT 2446
Db 483 GlyTyrMetThrGlnGlnLeuMetAsnLeuAlaGlyGlyAlaValValLeuAlaLeuGlu 502
QY 2447 GGAGCCATGACTGACCGCCATCTGTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2506
Db 503 GlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValAlaAlaLeuLeu 522
QY 2507 AGTGAAGCTGACCGCTTGGATGAGCAGTCTTGCAGCAAAAGCCCAACATCAACGCA 2566
Db 523 GlyAsnArgValAspProLeuSerGluGluGlyTrpLysGlnLysProAsnLeuAsnAla 542
QY 2567 GTGCCACGCTAGAAAGTCAATGAGTCCAGAGCAACACCTGGAGCTGTGTGCAGAG 2626
Db 543 IleArgSerLeuGluAlaValIleArgValHisSerLysTyrTrpGlyCysMetGlnArg 562
QY 2627 TTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2686
Db 563 LeuAlaSerCysProAspSerTrpValProArgValProGlyAlaAspLysGluGluVal 582
QY 2687 GAAAT-GTGAACGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2739
Db 583 GluAlaValThrAlaLeuAlaSerLeuSerValGlyIleLeuAlaGluAsp----- 599
QY 2740 CGCGCAGCCCGGACACACAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2793
Db 600 -----ArgProSerGluGlnLeuValGluGluGlu 609

RESULT 5
Q9NW41 PRELIMINARY; PRT; 614 AA.
ID Q9NW41;
AC Q9NW41;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA FLJ10328 fis, clone NT2RM2000588, weakly similar to histone
DE deacetylase HDAL.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001190; BAA91545.1;
DR InterPro; IPR000286; His_deacetylase.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PRO1270; HDASUPER.
SQ SEQUENCE 614 AA; 66173 MW; 36C82D314014F211 CRC64;

```

```

Alignment Scores:
Pred. No.: 7, 41e-89 Length: 614
Score: 1607.50 Matches: 343
Percent Similarity: 61.32% Conservative: 85
Best Local Similarity: 49.14% Mismatches: 131
Query Match: 30.07% Indels: 140
DB: 4 Gaps: 14

```

US-09-502-945-2 (1-2885) x Q9NW41 (1-614)

AC Q9H7L0;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE FLJ00062 protein (Fragment).
 GN FLJ00062.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SPLEEN;
 RA Ohara O., Nagase T., Kikuno R., Okumura K.;
 RT "the nucleotide sequence of a long cDNA clone isolated from human
 spleen.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK024469; BAB15759.1;
 DR InterPro; IPR000286; His_deacetylase.
 DR Pfam; PF00850; Hist_deacetyl; 1.
 DR PRINTS; PR01270; HDASUPER.
 FT NON_TER
 SO SEQUENCE 484 AA; 52327 MW; 1B1B0B363436A800 CRC64;

Alignment Scores:
 Pred. No.: 1-23e-83 Length: 484
 Score: 1520.50 Matches: 295
 Percent Similarity: 74.90% Conservative: 69
 Best Local Similarity: 60.70% Mismatches: 100
 Query Match: 28.44% Indels: 23
 DB: 4 Gaps: 5

US-09-502-945-2 (1-2885) x Q9H7L0 (1-484)

QY 1357 GGTGTACCAAGCCGCTCAGCTGGCCACTGTGCCACCACAGCCCTGGCGGTACCCA 1416
 Db 10 GlyValArgGlyValTyProProGlySerMetTrpValProAla-----ValAla 26
 QY 1417 ATCTCTCCCTGCTGC-----CCTGGGGGCGATGAAGAACCCTCCAGCAAC 1464
 Db 27 ValLeuAlaCysSerLeuGlnProArgProTrp---GlyValArgThr---ProTrpValPr 45
 QY 1465 CGTCAAGCACTCTTCCACCACAGTGTGGTCTAGGACACCTTCATGCTAAAGCACAGTG 1524
 Db 45 oAlaLeuThrLeuAlaProAlaGlyLeuIleTyAspSerValMetLeuLysHisGlnCy 65
 QY 1525 CATGTGGGGAACACACACGTGCACCTGAGCATGTGTCGGCGGATCCAGACGATCTGGTC 1584
 Db 65 sSerCysGlyAspAsnSerArgHisProGluHisAlaGlyArgIleGlnSerIleTrpSe 85
 QY 1585 CCGGCTCAGGACAGCCCTGCTTAGCAAGTGGAGCGGATCCGAGGTCCGAAAGCCAC 1644
 Db 85 rArgLeuGlnGluArgGlyLeuArgSerGlnCysGluCysLeuArgGlyArgLysAlaSe 105
 QY 1645 GCTAGATGAGATCCAGACAGTGCACCTGAATACCACACCTGCTGTATGGGACGATCC 1704
 Db 105 rLeuGluGluLeuGlnSerValHisSerGluArgHisValLeuLeuTyArgGlyThrAsnPr 125
 QY 1705 CCTCAAGCCGAGAGCTAGACAGCAAGAAGTTGCTGCTGCCATCGGCGAGAGATGTA 1764
 Db 125 oLeuSerArgLeuLysLeuAspAsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPh 145
 QY 1765 TGCTGTGCTGCTGTGGGGGATCGGGTGGACAGTGCACACCTGTGGATGAGATGCA 1824
 Db 145 eValMetLeuProCysGlyGlyValGlyValAspThrAspThrIleTrpAsnGluLeuHi 165
 QY 1825 CTCTCTCAGTGTGCGGATGCGAGTGGGCTGCTGCTGCTGGAGCTGGCTTCAAGGTGC 1884
 Db 165 sSerSerAsnAlaAlaArgTrpAlaAlaGlyArgValThrAspLeuAlaPhelLysValAl 185
 QY 1885 TGCAGAGAGCTCAAGAATGATTTGCCATCATCCGGCCCCAGGACACACGCCGAGGA 1944
 Db 185 aSerArgGluLeuLysAsnGlyPheAlaValValArgProProGlyHisHisAlaAspHi 205

QY 1945 ATCCACAGCCATGGGATTTCTTCAACTCTGTAGCATCACCAGCAAACTCCTACA 2004
 Db 205 sSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIleAlaCysArgGlnLeuGl 225
 QY 2005 GCAGAAAGTTGAACGTGGGCAAGTCTCATCGTGGACTGGGACATTCACCATGGCAATGG 2064
 Db 225 nGlnGlnSerLysAlaSerLysIleLeuIleValAspTrpAspValHisHisGlyAsnGl 245
 QY 2065 CACCCAGCAGCGCTTCTACAAATGACCCCTGTGCTCTACATCTCTCTGATCGCTATGA 2124
 Db 245 yThrGlnGlnThrPheTyArgGlnAspProSerValLeuTyIleSerLeuHisArgHisAs 265
 QY 2125 CAACGGGAATCTTCTCCAGGCTCTGGGGTCTGAAGAGTGTGGAGGACGACCGCGT 2184
 Db 265 pAspGlyAsnPhePheProGlySerGlyAlaValAspGluValGlyAlaGlySerGlyGl 285
 QY 2185 GGGGTACAATGTGAACGTGGCATGGACAGAGAGTGTGGACCCGCCCATTTGGAGACGTGA 2244
 Db 285 uGlyPheAsnValAsnValAlaTrpAlaGlyGlyLeuAspProProMetGlyAspProGl 305
 QY 2245 GTACCTTACAGCCTTCAGGACAGTGTGATGCCATTGCCACAGAGTCTCACCCTGATGT 2304
 Db 305 uTyLeuAlaAlaPheArgIleValValMetProIleAlaArgGluPheSerProaspLe 325
 QY 2305 GGTCTAGTCTCGCGGGTGTGATCTGTGAAGGACATCTGTCTCTCTGGTGGCTA 2364
 Db 325 uValLeuValSerAlaGlyPheAspAlaAlaGluGlyHisProAlaProLeuGlyGlyTy 345
 QY 2365 CTCTGTACCGCCGATGTTTGGCCACTGTGACGAGGACAGTGTGACCTGGCAGGGGG 2424
 Db 345 rHisValSerAlaLysCysPheGlyTyMetThrGlnGlnLeuMetAsnLeuAlaGlyGl 365
 QY 2425 CCGGTTGTGCTGGCCCTGGAGGAGGCGCATGACTGTACCGCCATCTGTGATGCCCTCTGA 2484
 Db 365 yAlaValValLeuAlaLeuGluGlyHisAspLeuThrAlaIleCysAspAlaSerGl 385
 QY 2485 AGCTTGTCTCGCTCTCTCAGTGTAAAGCTGTGAGCGCCCTTGGATGAGGCGAGTCTTGA 2544
 Db 385 uAlaCysValAlaAlaLeuLeuGlyAsnArgValAspProLeuSerGluGluGlyTrpLy 405
 QY 2545 GCAGAAAGCCACATCAACGCGAGTGGCCACCGCTAGAGAAGTCAATGAGATCCAGAGCAA 2604
 Db 405 sGlnLysProAsnLeuAsnAlaIleArgSerLeuGluAlaValIleArgValHisSerLy 425
 QY 2605 ACACGTGAGCTGTGTCAGAAGTGTGCGCGTGTGCTGGSCCGTCCCTCGAGGGGCCCA 2664
 Db 425 sTyTrpGlyCysMetGlnArgLeuAlaSerCysProaspSerTrpValProArgValPr 445
 QY 2665 AGCAGGTGAGACCGAAGAGCCGAAAT-GTGAACGCCATGGCTTGTCTGTGGTGGGG-- 2721
 Db 445 oGlyAlaAspLysGluGluValGluAlaValThrAlaLeuAlaSerLeuSerValGlyTrl 465
 QY 2722 ----GCCGACAGGCGCCAGCTGCGGACCGCCGGAACACACCCCGGCGGACAGGA 2777
 Db 465 eLeuAlaGluAsp-----ArgProSerGluGl 474
 QY 2778 GCCCATGGAGCAGGAG 2793
 Db 474 nLeuValGluGluGlu 479

RESULT 8
 Q96DY4
 ID Q96DY4 PRELIMINARY; PRT; 363 AA.
 AC Q96DY4;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Unknown (Protein for IMAGE:3867625) (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;


```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EYE;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013140; AAH13140.1; -
FT NON_TER 1
SQ SEQUENCE 363 AA; 40197 MW; 1EF788B16A0C6187 CRC64;

Alignment Scores:
Pred. No.: 7,14e-69 Length: 363
Score: 1275.00 Matches: 245
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.85% Indels: 0
DB: 4 Gaps: 0

US-09-502-945-2 (1-2885) x Q96DY4 (1-363)
QY 2 GAATTCCTCTGTCGAAGTCAAAGAGAGCCACACACAGGCGGCTCAACCATTCCTCCCA 61
DB 116 GluPheLeuLeuSerLysLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 135
QY 62 CAGCACCCCAATGCTGGGAGCCACCATGCTCTTTGGACAGAGTTCCCTCCCAAG 121
DB 136 GluHisProLysCysTrpGlyAlaHisHisAlaSerLeuAspGlnSerSerProGln 155
QY 122 AGCGGCGCCCTGGGAGCGCTCCTCTACAAACTGCCTTGGCTGGCGCTACGACAGT 181
DB 156 SerGlyProGlyThrProProSerTyrLysLeuProLeuProGlyProTyrAspSer 175
QY 182 CGAGACGACTTCCCTCCCAAAACAGAGCTCTGAACCAACTGAAAGTGCCTCAAGG 241
DB 176 ArgAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg 195
QY 242 CTAACACAGAGTGCTGAGCGGAGAGCAGTCCCTCTCGCTGCCAAGGATGGGACT 301
DB 196 LeuLysGlnLysValAlaGluArgSerSerProLeuLeuArgArgLysAspGlyThr 215
QY 302 GTTATTAGCACCTTTAAGAGAGAGCTGTGAGATCACAGTCCCGGCTGGCGCTCG 361
DB 216 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyAlaGlyProGlyAlaSer 235
QY 362 TCGGTGTGAACAGCGACCCGCTCCGGCCCGAGCTCTCCCAACAGCTCCCAACAGC 421
DB 236 SerValCysAsnSerAlaProGlySerGlyProSerSerProAsnSerSerHisSerThr 255
QY 422 ATCGTGAGAAATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCAG 481
DB 256 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGln 275
QY 482 CACGAGCCCTCCTCTGGACAGCTCCCAACAGCTTACAGCTCTACAGCTCTCCTCT 541
DB 276 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTyrThrSerProSer 295
QY 542 CTGCCCCAACATCTCCCTAGGCTGACGCGCACAGCTCACTGTCAACCACTCACACCTCACT 601
DB 296 LeuProAsnIleSerLeuGlyLeuGlnAlaThrValThrValThrAsnSerHisLeuThr 315
QY 602 GCCTCCCGGAGCTGTGCACACAGCAGGAGCGGAGAGGAGGAGGAGGAGGAGGAGGAG 661
DB 316 AlaSerProLysLeuSerThrGlnGlnGluAlaGluArgGlnAlaLeuGlnSerLeuArg 335
QY 662 CAGGCTGGCAGCTGACCGCAAGTTTATGAGCACATCTCTATTCTGGCTGCTGCTGCTG 721
DB 336 GlnGlyGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCysLeuLeu 736
QY 722 GCGGTGGCACTGGAG 736
DB 356 GlyValAlaLeuGlu 360

RESULT 9
Q91X14
```

```

ID Q91X14 PRELIMINARY; PRT; 221 AA.
AC Q91X14;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Type II histone deacetylase 5 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=HEART;
RA Potter G.B., Thompson C.C.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY038024; AAK71493.1; -
FT NON_TER 1
SQ SEQUENCE 221 AA; 23495 MW; 1467FBFE04770BA CRC64;

Alignment Scores:
Pred. No.: 3,03e-60 Length: 221
Score: 1131.00 Matches: 217
Percent Similarity: 98.19% Conservative: 0
Best Local Similarity: 98.19% Mismatches: 4
Query Match: 21.16% Indels: 0
DB: 11 Gaps: 0

US-09-502-945-2 (1-2885) x Q91X14 (1-221)
QY 11 TTGTCGAAGTCAAAGAGAGCCACACAGGCGGCTCAACCATTCCTCCCAACAGCACC 70
DB 1 LeuSerLysSerLysGluProThrProGlyGlyLeuAsnHisSerLeuProGlnHisPro 20
QY 71 AATGCTGGGAGAGCCACCATGCTCTTTGGACAGAGTTCCTCCCTCCCAAGCGGCCCC 130
DB 21 LysCysTrpGlyAlaHisHisAlaSerLeuAspGlnSerSerProGlnSerGlyPro 40
QY 131 CTGGGAGCCCTCCCTCTACAACTGCTTGGCTGGCGCTTACAGAGTGCAGACGAC 190
DB 41 ProGlyThrProProSerTyrLysLeuProLeuLeuGlyProTyrAspSerArgAspAsp 60
QY 191 TTCGCCCTCCGAAACAGCCCTCTGAACCCAACTTGAAGTCGGTCAAGGCTAAACAG 250
DB 61 PheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArgLeuLysGln 80
QY 251 AAGTGGCTGACGGAGAGCAGTCCCTCTCCCTCGCTCGCAAGATGGGACTCTTATTAGC 310
DB 81 LysValAlaGluArgSerSerProLeuLeuArgLysAspGlyThrValIleSer 100
QY 311 ACCTTTAAGAAGAGAGCTCTTGAGATCACAGTTCGCGGCTGGGCGCTCGTCCGTGTG 370
DB 101 ThrPheLysLysArgAlaValGluIleThrGlyThrGlyProGlyValSerValCys 120
QY 371 AACAGCGCCCGGCTCCCGGCCAGCTCTCCCAACAGCTCCCAACAGCACCACATCGCTGAG 430
DB 121 AsnSerAlaProGlySerGlyProSerSerProAsnSerSerHisSerThrIleAlaGlu 140
QY 431 AATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCAGCAGCAGCC 490
DB 141 AsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGlnHisArgAla 160
QY 491 CTCCCTCTGACAGCTCCCAACAGCTCTACAGCTCTACAGCTCTCTCTCTGCCCCAAC 550
DB 161 LeuProLeuAspSerSerProAsnGlnPheSerLeuTyrThrSerProSerLeuProAsn 180
QY 551 ATCTCCCTAGGGCTGACAGGCGCAGGTCACCTGTACCACTCACACCTCACTCACTCCCGG 610
DB 181 IleSerLeuGlyLeuGlnAlaThrValThrValThrAsnSerHisLeuThrAlaSerPro 200
QY 611 AAGCTGTGCACAGCAGGAGGCGGAGAGGCGGCTCCAGTCCCTCGCGGAGGGTGGC 670
DB 201 LysLeuSerThrGlnGlnGluAlaGluArgGlnAlaLeuGlnSerLeuArgGlnGlySer 220
```

```

QY 671 ACQ 673
Db 221 Thr 221

RESULT 10
O17323
ID O17323 PRELIMINARY; PRT; 796 AA.
AC O17323;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE C10E2.3 protein.
GN C10E2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Bonfield J., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wohldmann P., Sansone J.;
RT "The sequence of C. elegans cosmid C10E2.";
RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF026202; AAB71243.1; -.
DR InterPro; IPR000286; His_deacetylase.
DR DR InterPro; IPR000408; Reg_chromatase.
DR Pfam; PF00850; Hist_deacetylase; 1.
DR PRINTS; PR01270; HDASUPER.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
SQ SEQUENCE 796 AA; 86763 MW; 955619E75EBBCAB CRC64;

Alignment Scores:
Pred. No.: 2,84e-54 Length: 796
Score: 1034.00 Matches: 275
Percent Similarity: 43.91% Conservative: 96
Best Local Similarity: 32.54% Mismatches: 212
Query Match: 19.34% Indels: 262
DB: 5 Gaps: 23

US-09-502-945-2 (1-2885) x O17323 (1-796)
QY 122 AGCGGCCCCCTGGGAGCGCTCCCTCAAACTGCCT-----TTGCTGGGCCCTAC 175
Db 149 SerSerProHisPheGluPro-----TyrArgLeuProThrSerLeuAlaAsnAlaHis 166
QY 176 GAC-----AGTCGAGACAGACTTCCCTCCCAACAGCCTCTGAACCCCAACTTGAA 229
Db 167 AsnLeuGlnGlnAlaSerGluPheGlnLeuArgLysValAsnSerGluProAsnLeuLys 186
QY 230 GTGCGTTCAAGCTAAACAGAGAGTGGCTGAGCGGAGAGCAGTCCCTCTCTGCGTGC 289

```

```

Db 187 Met-----ArgIleArgAlaLysLeuLeuSerLysGlySerSerProValGlnHisVal 204
QY 290 AAGGATGGGACTGTTATTAGCACCTTTTAAGAAAGAGAGCTGTTGAGATCAGAGTGCCGGG 349
Db 205 GlnGlnAsnAsnSerGlnPheAsnPhe----- 213
QY 350 COTGGGGCGTCTGCTCCGTGTGTAAACAGCGCACCCGGTCCGGCCCGAGCTCTCCCAACAGC 409
Db 214 -----ThrHisProGlnLeuLysArgSerAspSerGluThr 225
QY 410 TCCACAGACCATCGCTGAGATGGCTTACTGGCTAGTCCCAACATCCCACTGAG 469
Db 226 SerGlnAsnValProLeuAsnPheMetGlnSerSerGlnThrAsnLeuProHisLeu 245
QY 470 ATGCTCCCTCTCAGCACCGAGCCCTCCTCTGGACAGCTCCCAACAGCAGTTCAGCTCTAC 529
Db 246 MetLeuPro----- 248
QY 530 ACGTCTCTCTCTGCGCCCAACATCTCCCTTAGGG----- 562
Db 249 ---SerProSerLeuProAsnLeuAlaAlaGlyAlaPheHisGlyLeuAsnLeuPro 267
QY 563 -----CTGCAGGCCAGCGTCACTGTCACTCACTCACTCACTCACTCACTCCCGG 610
Db 268 ValGlyGlnAspLeuAsnAlaPheMetAlaValAlaAsn-----LeuSerPro 283
QY 611 AAGCTGTGCACACAGCAGGAGCGCCGAGGAGCGCCCTCCAGCTCCCTCGCGGAGGTGGC 670
Db 284 PheLeuSerLeuProSerLeuLeuAsnLysLys-----LeuGluLeuGlyGly 299
QY 671 ACGGTGACCGGCAAGTTTCATGAGCACATCTCTATTCTTGTGCTGCTGCTGGCGGTGCA 730
Db 299 ----- 299
QY 731 CTGGAGGGCGAGCGGAGCGCCACCGGCATGCTCTCTGCTGCAGCATGTGCTGTGTG 790
Db 300 LeuThrAspGluGlyAspArgAsnGly----- 308
QY 791 GAGCAGGCGCGCAGCAGCAGCACCTCATTTGCTGTGCCACTCCACGGCAGTCCCACTA 850
Db 309 -----LeuIleGlySerSer----- 313
QY 851 GTGACGGGTGAACGTGTGGCCACCAGCATGGGAGCGTAGGCAAGCTCCCGCGCATCG 910
Db 313 ----- 313
QY 911 CCCCTGAGCGCGCACTCACTCTCACCCTGCGCGCAGAGTCCCGCGCCTCTGAGCAGGTG 970
Db 314 -----SerThrSerSerLeuAlaSerAsn 321
QY 971 GTCATGCAACAACAGCACCGAG-----CAGTTCTGTGGAGAAGCAGCAGCAGCAGTACAG 1027
Db 322 ValSerMetGlySerHisGlnTyrGlnSerLeuLeuLysGlnGlnIleArgAspLeuVal 341
QY 1028 CTGGCAAGATCTTCAACCAAGACAGGGGAGCTGCCAGGAGCCCAACCCACCTGAG 1087
Db 342 Leu----- 342
QY 1088 GAGACAGAGGAGGAGCTGACGAGCAGCAGGAGGTCTTGTGGGGGAGGAGCCCTGACC 1147
Db 342 ----- 342
QY 1148 ATGCCCCGGGAGGCTCCACAGAGTGCAGACAGCAGAGACCTGAGAGGAGGAGAC 1207
Db 343 -----ArgArgLysSerLeuValArgLysAsp 351
QY 1208 GAGGAAGAGGATGGGGAGGAGGAGGAGGATTCATCCAGGTTAAGGAGGAGGAGGAG 1267
Db 352 ProGlu----- 353
QY 1268 AGTGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1327

```

Db 354 -----GlyGluGlyLeuAlaGluLeuTyr 361
QY 1328 TCAGATGCCAACCGCTGCAACCTTTGACAGGTGTACCAAGGCCCTCAGCCCTGCGCCACT 1387
Db 362 AsnGlyLeuLeuProGlnAlaLysLeuGlnGlnAla ----- 375
QY 1388 GTGCCCAACCAAGCCCTGGCGGTACCCCAATCCCTGCTGCCCTGGGGGCATG--- 1444
Db 376 -----LeuAlaGluSerGlyPheLeu 383
QY 1445 ---AGAACCCCCAGACCAACCCGTCAGACACTCTTCACCAAGAGTGTGCTACGAC 1501
Db 384 AlaLysGlnGluPro -----ThrCysThrThrGlyLeuGlyTyrAsp 397
QY 1502 ACGTTTCATGCTAAAGCACCAGTCATGTCGGGAGACACACAGTCGACCCCTGAGCATGCT 1561
Db 398 GlnAlaMetValArgHisGluCysCysGlyAsnAsnAlaSerHisValGluAsnGly 417
QY 1562 GCGCGATCCAGAGCATCTGTCCTGGCTGCGAGGACAGGCGCTCTTAGCAAGTGGCAG 1621
Db 418 GlyArgIleGlnSerIleTrpSerLysLeuIleGluHisGlyHisValGlnLysCysGlu 437
QY 1622 CGATCCGAGGTGCGAAACCCAGCTAGATGATGATCCACAGTCGACTCTCAATACCAC 1681
Db 438 LysValThrAlaLysLysAlaSerLeuGluGlnLeuGlnLeuValHisSerGlnThrTyr 457
QY 1682 ACCTGCTCTATGGGACCACTGCCCTCAACCGGACAGAGCTAGACAGCAAGAAGTTGCTC 1741
Db 458 ThrThrPheAlaValSerProThrAlaCysLeuLysIleAspAlaAsnSerLeu--- 476
QY 1742 GGTCCATCAGCCAGAGATGATGCTGTGCTGCTGTCCTGTTGGGGGATCGGGGTGACAGT 1801
Db 477 ---ProLeu-----LysArgPheLeuGlnLeuProCysGlyIleGlyValAspSer 493
QY 1802 GACACGCTGTGAATGATGATGCTCTCCAGTGTGCTGCGATGCGATGCGAGTGGCTGCC 1861
Db 494 AspThrPheAsnAspAlaSerThrGlnThrAlaAlaArgLeuAlaGlyThrLeu 513
QY 1862 CTGGAGCTGGCTTCAAGTGGCTGCGAGGAGCTCAAGAATGGATTTCCCATCATCCGG 1921
Db 514 IleGluLeuSerSerGlnValAlaGluGlyArgLeuLysAsnGlyPheAlaCysIleArg 533
QY 1922 CCCCAGGACACACCCGAGGATCCACAGCATGGGATTTGCTTCTCACTCTGTA 1981
Db 534 ProProGlyHisHisAlaGluHisGluGlnAlaMetGlyPheCysPheAsnAsnVal 553
QY 1982 GCATCACCGCAAACTCTCAGCAGAGAG-----TTGACGTGGGCAAGTCTCTC 2035
Db 554 AlaValAlaValLysValLeuGlnThrLysTyrProAlaGlnCysAlaLysIleAlaIle 573
QY 2036 GTGGACTGGGACATTCACCATGGCAATGGCACCCAGCGCTTCTACAAATGACCCCTCT 2095
Db 574 IleAspTrpAspValHisHisGlyAsnGlyThrGlnLeuSerPheGluAsnAspProAsn 593
QY 2096 GTGCTCTACATCTCTGATGCTATGACACGGGAATCTTTCCAGGCTCTGGGGCT 2155
Db 594 ValLeuTyrMetSerLeuHisArgHisAspLysGlyAsnPhePheProGlyThrGlySer 613
QY 2156 CCTGAAGAGTGTGGTGAGGACCGCGGTGGGTACAATGTAACCTGCGCATGGACAGGA 2215
Db 614 ValThrGluValGlyLysAsnAspAlaLysGlyLeuThrValAsnValProPheSerGly 633
QY 2216 GGTGTGGACCCCCCATTTGGAGACGTGGAGTACCTTTACAGCCTTCAGGACAGTGTGTG 2275
Db 634 AspVal-----MetArgAspProGluTyrLeuAlaAlaTrpArgThrValIleGlu 650
QY 2276 CCAATGGCCACAGATCTCACTGATGTGGTCTAGTCTCCCGGGTTGATGCTGT 2335
Db 651 ProValMetAlaSerPheCysProAspPheIleValSerAlaGlyPheAspAlaCys 670
QY 2336 GAAGGACATCTGCTCTCTGGTGGCTACTCTGTCACCGCCAGAGTGTGTGGCCACTTG 2395
Db 671 HisGlyHisProAsnAlaLeuGlyGlyTyrGluValThrProGluMetPheGlyTyrMet 690

QY 2396 ACCAGGACGCTGATGACCTGGCAGGGGGCGGCTGTGCTGGCCCTGGAGGAGCCAT 2455
Db 691 ThrLysSerLeuLeuAsnTyrAlaSerGlyLysValValLeuAlaLeuGluGlyTyr 710
QY 2456 GACTTGACCGCCATCTGTGATGCTCTGAGCTGTGCTGCTGGCTGCTGCTC-----AGT 2509
Db 711 AspLysSerLysSerGluAlaGlnGlnCysValGlnAlaLeuIleGlyGluSer 730
QY 2510 GTAAAGCTGACGCTTGGATGAGCAGCTTTCAGCAAAAGCCCAACATCAACGCACTG 2569
Db 731 AspAspAlaGlyArgLeuSerSerValAlaLeuGluSerLeuProAsnProSerAlaVal 750
QY 2570 GCCACGCTAGAGAAA 2584
Db 751 GluThrLeuGlnLys 755
RESULT 11
Q96K29 PRELIMINARY; PRT; 191 AA.
AC Q96K29;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CDNA FLJ14828 fls, clone OVARC1000915, highly similar to Homo
DE sapiens histone deacetylase 5 mRNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARIAN CARCINOMA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamanoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027734; BAB55329.1; -.
DR InterPro; IPR000286; His_deacetylase.
DR Pfam; PF00850; Hist_deacetyl; 1.
SQ SEQUENCE 191 AA; 20877 MW; DCB6D45C0756A988 CRC64;
Alignment Scores:
Pred. No.: 4,74e-49 Length: 191
Score: 945.00 Matches: 179
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.68% Indels: 0
DB: 4 Gaps: 0
US-09-502-945-2 (1-2885) x Q96K29 (1-191)
QY 1508 ATGCTAAAGACACACAGTGTGCGGGAAACACACACGTCACCTGAGCATGCGCGG 1567
Db 1 MetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGluHisAlaGlyArg 20
QY 1568 ATCCAGAGCATCTGTCCCGGTGCGAGGACAGCGCTCTAGCAAGTGGCGGGATC 1627
Db 21 IleGlnSerIleTrpSerArgLeuGlnGluThrGlyLeuLeuSerLysCysGluArgIle 40
QY 1628 CAGAGTCCCAAGCCACACCTAGATGAGATCCAGACAGTGCACCTCTGATACACACCTG 1687
Db 41 ArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGlyThrHisThrLeu 60
QY 1688 CACTATGGACACAGTCCCTCAACCGGACAGAGCTAGACAGCAAGTGTCTGGTCCC 1747
Db 61 LeuTyrGlyThrSerProLeuAsnArgGlnLysLeuAspSerLysLeuLeuGlyPro 80


```
Score: 832.00 Matches: 206
Percent Similarity: 61.56% Conservative: 55
Best Local Similarity: 48.98% Mismatches: 123
Query Match: 15.56% Indels: 40
DB: 13 Gaps: 15

US-09-502-945-2 (1-2885) x Q9NGY4 (1-565)
QY 2 GAATTCCTCTTTCGAAGTCA-----AAGAGCCACACAGCGGGCTCAACCATGCC 55
Db 150 GluPheIleLeuSerLysAlaThrLysGluProLeuThrAsnGlyThrSerHisSer 169
QY 56 CTCACACAGACCCCAAAAGCTGG-----GGAGCCACCATCTCTTGGACACAGCT 109
Db 170 MetGlyArgHisProLysLeuIleThrAlaAlaHisHisThrSerLeuAspGlnSer 189
QY 110 TCCCTTCCCAAGAGGGCCCCCTGGGAGCTCTCCCTCTCTCTCTCTCTCTCTCTCTCT 169
Db 190 SerProPro-----ProSerGlyThrSerProThrThrLysCysProProGly 206
QY 170 CCTACGACAGTCGAGACGACTTCCCTCCGCAAAACAGCTCTCTGAACCAACTTGA 229
Db 207 -----AsnGlnAspPheProLeuArgLysThrAlaSerGluProAsnLeuLys 223
QY 230 GTGCGTCAAGGCTAAACAGAGGTGGCTGAGCGGAGAGCAGTCCCTCTCTCTCTCTCT 289
Db 224 ValArgSerArgLeuLysGlnLysValValGluArgSerSerProLeuLeuArgArg 243
QY 290 AGGATGGAGCTGTTATAGCACCTTTAGAGACGAGCTGTGAGATCAGAGTCCGCGG 349
Db 244 LysAspSerIleValSerSerTyrLysLysArgIlePheGluVal----- 259
QY 350 CCGTGGGCGTGCCTGTGTACACGACCGCGCTCCGCGCCAGCTCTCCCAACAGC 409
Db 260 ---AlaGluSerSerValSerSerProValSerGlyProSerSerProAsnAsn 278
QY 410 TCCACAGCACCATCTGAGATGGCTTTACTGGC-----TCAGTCCCAACATCCCC 463
Db 279 GlyProValAlaMetGluAlaGluHisGluThrProValLeuSerValAsnSerArgIle 298
QY 464 ACTGAGATGCTCCTCAGCACCGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 523
Db 299 GluAsnLeuValSerHisHisHisLeuValHisGluArgSerLeuSerLeuLeuAsn 318
QY 524 CTCTACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 583
Db 319 LeuThrThrSerProSerLeuProAsnIleThrLeuGlyLeuHisAlaThrAlaThr 337
QY 584 ACCAACTCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 643
Db 338 -----GlnLeuAsnThrSerSerSerLeuLysGluGlnGlnLysTyrAspProGln 354
QY 644 GCCCTCAGTCCCTGCGGAGGTGGGACGCTGACCGGCAAGTTCATGACCATCTCTCT 703
Db 355 Ala-----ProArgGlnGlyValSerMetAlaGlyGlnTyr-----AlaGlyGly 369
QY 704 ATTCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 763
Db 370 IleProThrSerSerAsnHisValSerLeuGlu-----GlyLysAlaAsnSerHisGln 387
QY 764 TCCCTGCTGACGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 823
Db 388 AlaIleLeuGlnHisLeuLeuLysGluGlnMetArgGlnGlnLysIleLeuAlaSer 407
QY 824 -----GTGGCACTCCAGGGCAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 877
Db 408 GlyGlyThrProValLeuHisGlnSerProLeuAlaAlaLysAspArgValSerProAla 427
QY 878 ATGGGAGGTAGCAGCTCCCGCGCATCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 937
Db 428 GlyArgValAlaHisLysLeuProArgHisArgProLeuHisArgThrGlnSerAlaPro 447
QY 938 CTGGCGCAGAGTCCCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 997

LeuProGlnSer-----ThrLeuAlaGlnLeuValIleGlnGlnGlnHisGlnGlnPhe 465
CTGGAGAGCAGAG-----CAGCAGCAGCTACAGCTGGCGAGATCCTCACCACAGACA 1051
LeuGluLysGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 485
GGGAGCTGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1111
IleGluGlnLeuArgGlnProGluGlyHisLeuGluGluAlaGluGluAspLeu----- 503
CAGCAGAGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1171
HisGlyAspAsnLeuMetGlnGlnLysSerSerSerIleAspAsnThrArgSerTyrSer 523
AGTGAGACACACAGGAA-----GACCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 1219
SerThrAspLeuArgThrGlyProPheGlySerValLysValLysGluGluProAsp 543
GGGAGGAGGAG 1231
SerGluAsnGlu 547

RESULT 14
Q9NZS3
ID Q9NZS3 PRELIMINARY; PRT; 1066 AA.
AC Q9NZS3;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE GR AF-1 specific histone deacetylase (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Govindan M.V., Warriar N.;
RT "Human GR AF-1 specific histone deacetylase.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF174495; AAF36540.1;
DR InterPro; IPR000286; His_deacetylase.
DR InterPro; IPR001607; ZnF_UBP.
DR Pfam; PF00850; Hist_deacetyl; 2.
DR Pfam; PF02148; zf-UBP; 1.
DR PRINTS; PR01270; HDASUPER.
DR SMART; SM00290; ZnF_UBP; 1.
FT NON_TER 1
SQ SEQUENCE 1066 AA; 114284 MW; 7001E35AAC093346 CRC64;

Alignment Scores:
Pred. No.: 2,17e-39 Length: 1066
Score: 787.50 Matches: 228
Percent Similarity: 46.40% Conservative: 88
Best Local Similarity: 33.48% Mismatches: 236
Query Match: 14.73% Indels: 129
DB: 4 Gaps: 24

US-09-502-945-2 (1-2885) x Q9NZS3 (1-1066)
QY 898 CCGCGCGATCGGCCCTCTGAGCGGCGACTCAGCTCCTCAGCTGCCGCGAGTCCCCAGGC 957
Db 180 ProValAlaLeuGluPheGlnProGlnLeuValLeuValAlaAlaGly----- 195
QY 958 CTGCGAGCAGCTGCTCATGCAACACAGCAGCAGCAGCTCTCTGGAGAA-----GCAGAA 1011
Db 196 -----PheAspAlaLeuGlnGlyAspProLysGlyGlu 206
QY 1012 GCAGCAGCAGCTACAGCTGGGCAAGATCTCTCACCACAGCAGGAGGAGTCCCCAGGC 1071
Db 207 MetAlaAlaThrProAla-----GlyPheAlaGlnLeuThr 218
```

```
QY 1072 CACCACCCACCTGCTGAGGAGACAGAGGAGCTGACGGAGCAGCAGGAGCTTGTCTGGG 1131
    |||
Db 219 His -----LeuLeuMetGlyLeuAlaGly 226
QY 1132 GGAGGGAGCCCTGACCATGCCCGGGAGGCTCCACAGAGAGTGAGACACAGGAAGA 1191
    |||||
Db 227 GlyLysLeuLeuSerLeuGluGlyGlyTyraAsnLeu ----- 239
QY 1192 CCTGGAGGAGGAGCAGGAGGAGGATGGGAGGAGGAGGATTCATCCAGTTAA 1251
    |||
Db 240 -----ArgAlaLeuAlaGluGlyValSerAlaSerLeuHis-ThrLeuLe 254
QY 1252 GGACGAGGAGGCGAGAGTGCTGTGAGGAGGCGCCAGCTTGAGGAGGCTGTGTCTGG 1311
    |||||
Db 254 u-----GlyAspProGlyProMetLeuGluSerProGlyAlaPr 267
QY 1312 ATACAAAAACTGTTCTACAGATGCCAA-----CCGCTGCAACCTTTGAGGTGTACCAAGC 1368
    |||
Db 267 oCysArgSerAlaGlnAlaSerValSerCysAlaLeuGluAlaLeuGluProPheTrpGl 287
QY 1369 GCCCTC-----AGCTGCGCCATGTGCCCCACCAAGCCCTGGCGGTACCCAAATCCTC 1422
    |||
Db 287 uValLeuValArgSerThrGluThrValGluArgGlyHisMetGlyGluAspAsnValGl 307
QY 1423 CCCTGCTGCCCTGGGGCATGAAGAACCCCGACACCAACCCGTCACGAC----- 1474
    |||
Db 307 uGluSerGluGluGlyProTrpGluProProValLeuProIleLeuThrTrpProVa 327
QY 1475 -----CTCTTCCACCAAGTGTGTACGACACGTTTCATGCTAAAGCACAGTGTCATGTG 1530
    |||
Db 327 lLeuThrSerArgThrGlyLeuValTyraAspGlnAsnMetMetAsnHis---CysAsnLe 346
QY 1531 CGGGAACACACAGTCGACCCCTGAGCATGCTGCCGATCCAGAGCATGTGTCCCGGCT 1590
    |||||
Db 346 uTrpAspSerHis---HisProGluValProGlnArgIleLeuArgIleMetCysArgLe 365
QY 1591 GCAGGAGACAGGCTGCTTAGCAAGTGCAGGAGTCCGAGGTCCGAAAGCCACGCTAGA 1650
    |||||
Db 365 uGluGluLeuGlyLeuAlaGlyArgCysLeuThrLeuThrProArgProAlaThrGluAl 385
QY 1651 TGAGATCCAGACAGTGCATCT---GAATACACACCCCTGCTATGGACCATGCCCT 1707
    |||||
Db 385 aGluLeuLeuThrCysHisSerAlaGluTyraValGlyHisLeuArgAlaThrGluLysMe 405
QY 1708 CAACCGGACAGACGTAGACAGCAAGATGTCTGCTCCATCCATCAGCCAGCATGTATGC 1767
    |||||
Db 405 tLysThrArgGluLeuHisArgGlu----- 413
QY 1768 TGTGCTGCCTTGTGGGGCATCGGGTGGACGTGACACCGTGTGGAATGAGATGCATC 1827
    |||
Db 414 -----SerSerAsnPheAspSerIleTyrlleCysProSerTh 426
QY 1828 CTCCAGTGTGTGCGCATGGCAGTGGCTGCCCTGCTGGAGCTGGCCCTCAAGTGGCTGC 1887
    |||
Db 426 rPheAlaCysAlaGlnLeuAlaThrGlyAlaAlaCysArgLeuValGluAlaValLeuSe 446
QY 1888 AGGAGAGCTCAAGATGATTTCCATCATCGGCCCGCCAGACACCCAGCCGAGGAATC 1947
    |||||
Db 446 rGlyGluValLeuAsnGlyAlaAlaValValArgProProGlyHisHisAlaGluGlnAs 466
QY 1948 CACAGCATGGGATTCCTCTTCAACTCTGTAGCATCCAGCAAACTCCTACAGCA 2007
    |||
Db 466 pAlaAlaCysGlyPheCysPhePheAsnSerValAlaValAlaAlaArgHisAlaGlnTh 486
QY 2008 G---ANGTTGACGTGGGCAAGTCTCATCGTGGACTGGGACATTCACCATGGCAATGG 2064
    |||
Db 486 rIleSerGlyHisAlaLeuAlaArgIleLeuValAspTrpAspValHisHisGlyAsnGl 506
QY 2065 CACCAGCAGGCGTCTCAATGATCCCTCTGTGCTCTACATCTCTCTCATCGCTATGA 2124
    |||||
Db 506 yThrGlnHisMetPheGluAspAspProSerValLeuTyraValSerLeuHisArgTyraS 526
QY 2125 CAACGGGAACCTCTTTCCA-----GGCTCTGGGGCTCTGAGAGGTGGTGGAGGACC 2178
```

```
Db 526 pHisGlyThrPhePheProMetGlyAspGluGlyAlaSerSerGlnIleGlyArgAlaAl 546
QY 2179 AGCGTGGGTACAATGTGAACCTGGCATGGCAGGAGGTGTGGACCCCGCCATTTGGAGA 2238
    |||
Db 546 aGlyThrGlyPheThrValAsnValAlaTrpAsnGly-----ProArgMetGlyAs 563
QY 2239 CGTGGAGTACCTTACAGCCTTCAGGACAGTGTGGTATGCCCATTCGCCAGAGTTCACCC 2298
    |||||
Db 563 pAlaAspTyrrLeuAlaAlaTrpHisArgLeuValLeuProIleAlaTyrrGluPheAsnPr 583
QY 2299 TGATGTGGTCTAGTCTCCCGGGTTTGTATGCTGTGTGAGGACATCTGTCTCTCTGGG 2358
    |||||
Db 583 oGluLeuValLeuValSerAlaGlyPheAspAlaAlaArgGly-----AspProLeuGl 601
QY 2359 TGCTACTCTGTCCAGCCAGATGTTTGGCCACTTGACCGACGACGTGATCACCCCTGGC 2418
    |||||
Db 601 yGlyCysGlnValSerProGluGlyTyrrAlaHisLeuThrHisLeuLeuMetGlyLeuAl 621
QY 2419 AGGGCCCGGGTGTCTGCTGCCCTGGAGGAGGCCATGACTTGACCGCCATCTGTGATGC 2478
    |||||
Db 621 aSerGlyArgIleIleLeuIleLeuGluGlyGlyTyrrAsnLeuThrSerIleSerGluSe 641
QY 2479 CTCTGAAGCTTGTCTCGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 2538
    |||||
Db 641 rMetaAlaAlaCysThrArgSerLeuLeuGly-----AspProProProLeuLeuTh 658
QY 2539 CTTCGAGCAAAAGCCCAACATCAAGCAGTGGCCGCTAGAGAAAGTCAATCATCGATCCA 2598
    |||
Db 658 rLeuProArgProProLeuSerGlyAlaLeuAlaSerIleThrGluThrIleGlnValHi 678
QY 2599 GAGCAAACTGAGAGTGTGTGCAGAAAGTTCGCCGCTGCTGCGCGCTGCTGCGAGG 2658
    |||||
Db 678 sAlaArgTyrrTrpCysSerLeu-Arg-----ValMetLysValGluAspArgGluG 695
QY 2659 GCGCC-----AACGAGTGTAGACCGCAAGACCGCAAGACCGAAATGT- 2693
    |||||
Db 695 lProSerSerSerLysLeuValThrLysLysAlaProGlnProAlaLysProArgLeuA 715
QY 2694 -----GAACCCATGGCTTGTCTGTTGGTGGGCGCA---ACAG 2730
    |||||
Db 715 laGluArgMetThrThrArgGluLysLysValLeuGluAlaGlyMetGlyLysValThrS 735
QY 2731 GCCCAAGTGGCG-----AGCCCGGA----- 2753
    |||
Db 735 eAlaSerPheGlyGluGluSerThrProGlyGlnThrAsnSerGluThrAlaValAla 755
QY 2754 -----ACACAGCCCGCGCGCAGGAGCCCATGGAGCAGGAGCCTGCCCTGTGACG 2807
    |||||
Db 755 laLeuThrGlnAspGlnProSerGluAlaAlaThrGlyGlyAlaThrLeuAlaGlnThr 774
RESULT 15
Q9NSW6
ID Q9NSW6 PRELIMINARY; PRT; 1063 AA.
AC Q9NSW6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 114.4 kDa protein.
GN DKFZP566E044.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=KIDNEY;
RA Bloembergen H., Boecher M., Brandt P., Mewes H.W., Gassenhuber J.,
RA Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137696; CAB70878.1;
DR InterPro; IPR000286; His_deacetylase.
DR InterPro; IPR001607; ZnF_UBP.
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 08:38:27 ; Search time 31.7038 Seconds
(without alignments)
10910.938 Million cell updates/sec

Title: US-09-502-945-3
Perfect score: 2140
Sequence: 1 ggcgtgtaagtactgcga.....gaagaagaagaaggaaaaa 1298

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09502945/runat_14032003.101057_19087/app-query.fasta_1.10979
-DB=A_Geneseq.101002 -QMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09502945.ecgn.1.1.365 -runat_14032003.101057_19087 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1047	48.9	236	20	Colon cancer assoc
2	677	31.6	157	22	Human protein sequ
3	401	18.7	434	18	Antibiotic potenti
4	401	18.7	434	21	(Lys)434 protein s
5	365	17.1	554	22	Drosophila melanog
6	354.5	16.6	808	22	Novel human diagno
7	326	15.2	71	21	Human secreted pro
8	325.5	15.2	215	22	Novel human diagno
9	325.5	15.2	215	22	Novel human diagno
10	314.5	14.7	182	22	Novel human diagno
11	313	14.6	611	20	T. gondii immunoge
12	313	14.6	611	22	T. gondii immunoge
13	312	14.6	274	22	Drosophila melanog
14	309.5	14.5	771	22	Drosophila melanog
15	309	14.4	405	22	Novel human diagno
16	307.5	14.4	1178	18	Mannose-1-phosphat
17	304	14.2	436	22	Novel human diagno
18	298	13.9	1898	20	A human trichohyal
19	290.5	13.6	389	22	Novel human diagno
20	286.5	13.4	328	22	Novel human diagno
21	285.5	13.3	573	22	Novel human diagno
22	281	13.1	843	22	Human protein SEQ
23	280	13.1	204	22	Novel human diagno
24	277.5	13.0	191	22	Novel human diagno
25	276.5	12.9	695	22	Human protein SEQ
26	273	12.8	189	22	Novel human diagno
27	266.5	12.5	347	22	Human polypeptide
28	266	12.4	140	22	Novel human diagno
29	266	12.4	624	23	Human RNA metaboli
30	265	12.4	100	21	Peptide #8 used in
31	265	12.4	100	21	Amino acid polymer
32	265	12.4	100	22	Poly-L-Lysine used
33	265	12.4	100	22	Nucleic acid trans
34	264	12.3	146	22	Novel human diagno
35	264	12.3	1098	22	Novel human diagno
36	264	12.3	1333	22	Novel human diagno
37	264	12.3	1430	22	Drosophila melanog
38	263	12.3	361	20	Drosophila melanog
39	261.5	12.2	194	22	Biorhythm marker p
40	261.5	12.2	540	22	Novel human diagno
41	261	12.2	1074	22	Human Ser/Arg rich
42	261	12.2	1507	21	Novel human diagno
43	259	12.1	412	17	Plasmodium yoelii
44	259	12.1	735	22	Human thyrotropin
45	258	12.1	191	22	Novel human diagno

ALIGNMENTS

RESULT 1
AA07093
ID RAY07093 standard; Protein; 236 AA.
XX
AC RAY07093;
XX
DT 02-JUL-1999 (first entry)
DE
DE Colon cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.


```
XX PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Disclosure; Page 655; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 236 AA;
Alignment Scores:
Pred. No.: 9,06e-92 Length: 236
Score: 1047.00 Matches: 231
Percent Similarity: 95.85% Conservative: 0
Best Local Similarity: 95.85% Mismatches: 4
Query Match: 48.93% Indels: 8
DB: 20 Gaps: 5
US-09-502-945-3 (1-1298) x ANY07093 (1-236)
QY 2 GCTGCTGAATGACTGCGAACCGGCTTGCAGAGAGCGCTTCTGGCTTTGAGCCANCAAGAA 61
Db 1 AlaAlaGluMetThrAlaAsnArgLeuAlaGluSerLeuLeuAlaLeuSer---GlnGlu 19
QY 62 GAACACTGCGGATTTGCCAAAGACTACCTCTTGAGTGAGAGTGAAGATGAGGGGACAAT 121
Db 20 GluLeuAlaAspLeuProIysAspTyrLeuLeuSerGluSerGluAspGluGlyAspAsn 39
QY 122 GATGGAGAGAGAACATCAACATCTCTGGAGACCAATCAGTTCCTGATGGAAAGAAAT 181
Db 40 AspGlyGluArgLysHis---LysLeuLeuGluAlaAlaSerSerLeuAspGlyLysAsn 58
QY 182 AGGCGGAAATTTGGCTGANAGGCTGTGAGGCTAGTCTGAAGGTGTGAGAGTTCAATGTGAGT 241
Db 59 ArgArgLysLeuAla---ArgSerGluAlaSerLeuLysValSerGluPheAsnValSer 77
QY 242 TCTGAAGGATCAGAGAAAGCTGGTCCCTGAGATCTGCTTGGAGCTGTAAACTTCA 301
Db 78 SerGlyGlySerGlyGluLysLeuValLeuAlaAspLeuLeuGluProValLysThrSer 97
QY 302 TCTTCTTGGCCACTGTGAAAGCAACTCAGTAGCTCANATCAAGAANAACAGTGGAG 361
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

```
Db 98 SerSerLeuAlaThrValLysLysGlnLeuSerArgValSerLys-----ThrValGlu 115
QY 362 TTACCTCTGAACAAGAGAGATTGAACGGATCCACAGAGAATA-GCATTCAATAAAGC 420
Db 116 LeuProLeuAsnLysGluGluLeuArgIleHisArgGluIleAlaPheAsnLysThr 135
QY 421 CACAAGTCTCTCAAAATGGGACCTGTC---GTCCTGAAGAACCAGCGCAGAGCAG 476
Db 136 HisLysSerSerProAsnGlyThrLeuSerValLeuLysAsnArgGlnAlaGluGln 155
QY 477 CTGGTTTTTCCCTGGAGAAAGAGAGCGACCATTTGCTCCATTGAACATGTGCTCAGT 536
Db 156 LeuValPheProLeuGluLysGluGluProAlaIleAlaProIleGluHisValLeuSer 175
QY 537 GGCTGGAAGGCAAGAACTCCCTCGAGCAGGAGAAATTTTCAACCTCCTCCATAAGAACAAG 596
Db 176 GlyTrpLysAlaArgThrProLeuGluGlnGluIlePheAsnLeuLeuHisLysAsnLys 195
QY 597 CAGCAGTGCAGACCCCTTTACTGACCCCTGTGGAAAGGCTCTCTCCGAGCCATGAGC 656
Db 196 GlnProValThrAspProLeuLeuThrProValGluLysAlaSerLeuArgAlaMetSer 215
QY 657 CTAGAAGAGGCAAGATCGAGCAGCAGCTTCAGAGGGCTCGGGCTCTGCAGTCTCTAC 716
Db 216 LeuGluGluAlaLysMetArgAlaGluLeuGlnArgAlaArgAlaLeuGlnSerTyr 235
QY 717 TAT 719
Db 236 Tyr 236
RESULT 2
AAAM25909
ID AAAM25909 standard; Protein; 157. AA.
... XX
AC AAAM25909;
XX
XX 16-OCT-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:1424.
XX
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
XX antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
XX antibacterial; endocrine; cardiant; central nervous system; virucide;
XX anti-HIV; fungicide; antitumagen; cardiovascular; antianaemic; anaemia;
XX antiaggregant; haemostatic; vulnerary; antileuc; osteopathic; eczema;
XX dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
XX neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
XX immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
XX antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
XX cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
XX genetic disease; haematopoietic disorder; platelet disorder; asthma;
XX thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
XX allergic rhinitis; diabetes; multiple sclerosis; depression;
XX Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
XX neurological disorder.
XX
XX Homo sapiens.
XX
XX WO200153455-A2.
XX
XX 26-JUL-2001.
PD
XX
XX 22-DEC-2000; 2000WO-US35017.
PF
XX
XX 23-DEC-1999; 99US-0471275.
PR
XX 21-JAN-2000; 2000US-0488725.
PR
XX 25-APR-2000; 2000US-0552317.
PR
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX WPI; 2001-457603/49.
DR
```

DR N-PSDB; AAH99850.
 XX Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 XX
 PS Claim 20; Page 289; 1217pp; English.
 XX
 CC AAH99166 to AAH99904 encode the human proteins given in AAH25225 to
 CC AAH25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardian;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;
 CC antiulcer; osteopathic; dermatological; antiallergic; antisthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 XX

SQ Sequence 157 AA;

Alignment Scores:
 Pred. No.: 2,44e-56 Length: 157
 Score: 677.00 Matches: 150
 Percent Similarity: 95.54% Conservative: 0
 Best Local Similarity: 95.54% Mismatches: 7
 Query Match: 31.64% Indels: 2
 DB: 22 Gaps: 0

US-09-502-945-3 (1-1298) x AAH25909 (1-157)

QY 11 ATGACTGCGAACCGCTTGCAGAGACCTTCTGGCTTGAGCCANCAAGAGAACTAGCG 70
 Db 1 MetThrAlaAsnArgLeuAlaGluSerLeuLeuAlaLeuSerGlnGlnGluLeuAla 20
 QY 71 GATTTGCCAAAGACTACTCTTGAGTGAGAGTGAAGATGAGGGGACAATGATGAGAG 130
 Db 21 AspLeuProLysAspTyrLeuLeuSerGluSerGluAspGluGlyAspAsnAspGlyGlu 40
 QY 131 AGAAAGCATCNAAGCTTCTGGAACCAATCAGTTCCTTGGATGGAAGAANTAGCGGAAA 190
 Db 41 ArgLysHisGlnLysLeuLeuGluAlaIleSerSerLeuAspGlyLysAsnArgLys 60
 QY 191 TTGGCTGANAGGCTCTGAGGCTAGTCTGAAGGTGTGAGAGTTCAGTTCAGTCTCTGAAGGA 250
 Db 61 LeuAlaGluArgSerGluAlaSerLeuLysValSerGluPheAsnValSerSerGluGly 80
 QY 251 TCAGGAGAAAAGCTGGTCCCTTGCAGATCTGCTTGAGCCTGTAAACCTTCATCTCTTTG 310
 Db 81 SerGlyGluLysLeuValLeuAlaAspLeuLeuGluProValLysThrSerSerLeu 100
 QY 311 GCCACTGTGAAGAACAACCTAGTAGAGTCANATCAAGAANACAGTGGAGTTACCTCTG 370
 Db 101 AlaThrValLysLysGlnLeuSerArgValLysSerLysLysThrValGluLeuProLeu 120
 QY 371 AACAAAGAAGAGATTGAACGATCCACAGAGAATA-GCATTCATAAACAACCA-CAGATC 428
 Db 121 AsnLysGluGluIleGluArgIleHisArgGluValAlaPheAsnLysThrAlaGlnVal 140
 QY 429 CTCTCCAAATGGGACCTGTGCTCTGAAGAACCGGACGAGGACAGCAGCTG 479
 Db 141 LeuSerLysTrpAspProValValLeuLysAsnArgGlnAlaGluGlnLeu 157

RESULT 3

AAW21591
 ID AAW21591 standard; peptide; 434 AA.
 XX
 AC AAW21591;
 XX
 XX 26-AUG-1997 (first entry)
 DT
 XX
 DE Antibiotic potentiating peptide #3.
 XX
 KW Potentiate; antibiotic; microbial infection; lipopolysaccharide;
 KW permeability; outer bacterial membrane.
 XX
 OS Synthetic.
 OS
 PN WO9638163-A1.
 PN
 XX
 PD 05-DEC-1996.
 PD
 XX
 PF 29-MAY-1996; 96WO-EP02313.
 PF
 XX
 PR 31-MAY-1995; 95US-0456112.
 PR
 XX
 PA (BIOS-) BIOSYNTH SRL.
 PA
 XX
 PI Porro M, Varra M;
 PI
 XX
 DR WPI; 1997-034095/03.
 DR
 XX
 PT Potentiating activity of antibiotic with peptide contg. cationic
 PT amino acid sequence - reduces dose of antibiotic required
 XX
 PS Disclosure; Page 7; 37pp; English.
 XX
 CC The sequences given in AAW21589-633 represent peptides which act to
 CC potentiate the activity of an antibiotic when they are co-administered
 CC with the antibiotic. Compositions containing these peptides are used
 CC to treat or prevent microbial infections. These peptides bind to
 CC lipopolysaccharide on the bacteria so may increase permeability of
 CC the outer bacterial membrane to the antibiotic, allowing a reduction
 CC in the dose of antibiotic required by 10-90% of the normal dose for
 CC in vivo or in vitro application. Any toxic side effects are
 CC correspondingly reduced.
 XX
 SQ Sequence 434 AA;
 Alignment Scores:
 Pred. No.: 9.93e-30 Length: 434
 Score: 401.00 Matches: 79
 Percent Similarity: 73.33% Conservative: 86
 Best Local Similarity: 35.11% Mismatches: 56
 Query Match: 18.74% Indels: 4
 DB: 18 Gaps: 0
 US-09-502-945-3 (1-1298) x AAW21591 (1-434)
 QY 630 GAAAGGCTCTCTCCGAGCATGAGCTAGAGAGGCAAGATGCGACGAGAGAGCTT 689
 Db 1 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 20
 QY 690 CAGAGGCTCGGGCTCTGCAGTCTCTACTATGAGCCAGGCTCGAAGAGAGAGAAATC 749
 Db 21 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
 QY 750 NAAAGTTAAAGTATCAAAAGTCGTGAAGAGGAAAGCCCAAGAAAGCCCTAAAGAG 809
 Db 41 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 60
 QY 810 TTTGAGCAGCTCGGAGGTTAATCCAGCTGCCGACTAGAACCAACCAAGAGAGAA 869
 Db 61 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
 QY 870 AGAAGGAGGAGGAGAAAGAAAGAACAGAGAGAAAGAAAGAGGAGGAGAGAGA 929

DR WPI; 2000-500381/45.
DR N-PSDB; AAC02146.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 6221; 71pp + CD-ROM; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 71 AA;
Alignment Scores:
Pred. No.: 8.72e-23 Length: 71
Score: 326.00 Matches: 68
Percent Similarity: 95.77% Conservative: 0
Best Local Similarity: 95.77% Mismatches: 3
Query Match: 15.23% Indels: 0
DB: 21 Gaps: 0
US-09-502-945-3 (1-1298) x AAG02140 (1-71)
QY 11 ATGACTGCGAACCGCTTGAGAGAGCTTCTGGCTTTGACCCAGGAGACTAGCG 70
Db 1 MetThrAlaAsnArgLeuAlaGluSerLeuLeuAlaLeuSerGlnGlnGluLeuAla 20
QY 71 GATTGTCGCAAAAGACTACCTCTTGAGTGAGGTGAAGATGAGGGGACAATGATGGAG 130
Db 21 AspLeuProLysAspTyrLeuLeuSerGluSerGluAspGluGlyAspAsnAspGlyGlu 40
QY 131 AGAAGCATCNAAGCTCTGGAGCAATCAGTCCCTTGATGGAAGATAGCGGAAA 190
Db 41 ArgLysHisGlnLysLeuLeuGluAlaIleSerSerLeuAspGlyLysAsnArgArgLys 60
QY 191 TTGGCTGANAAGTCTGAGGCTAGTCTGAAGGTG 223
Db 61 LeuAlaGluArgSerGluAlaSerLeuLysVal 71
RESULT 8
ID ABG26522 standard; Protein; 152 AA.
XX
AC ABG26522;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #26513.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Dmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS90709.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 56881; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC imaging of sites expressing (II). (I) and (II) are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 152 AA;
Alignment Scores:
Pred. No.: 1.24e-22 Length: 152
Score: 325.50 Matches: 66
Percent Similarity: 75.89% Conservative: 41
Best Local Similarity: 46.81% Mismatches: 24
Query Match: 15.21% Indels: 10
DB: 22 Gaps: 4
US-09-502-945-3 (1-1298) x ABG26522 (1-152)
QY 842 CGCAGTGAAGAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
Db 19 ArgArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluGly 38
QY 902 GAAGAAGAAGAAG 961
Db 39 GluGlu-----GluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 54
QY 962 AGAAG 1021
Db 55 ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 74
QY 1022 GAGGAA---GAATAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1078
Db 75 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 93
QY 1079 GAAGGAAGAAG 1138
Db 94 GluGlyArgArgArg-ArgArgArgArgArgArgArgArgArgArgArgArgArgArg 113
QY 1139 GGAACNAG 1198
Db 113 g-----ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 130


```
Db 130 AsnAspGlyAspArgGluGlyAspGluLysGluThrArgAspValGluAspGluGlyGlu 149
      :::::::::::::: ::::: ||| ::::: |||
QY 155 -----GCAATCACTCCCTGATGGAAGAAATAGCGGAATTTGGCTGAN 199
      ::::: ||| ||| |||
Db 150 ThrArgArgSerSerPheAlaGluGlnThrGlyAsnGluArgThrGluMetArgThr 169
      ::::: ||| ||| |||
QY 200 AGG----- 202
      |||
Db 170 ArgHisGlyClyAspGluGlyTrpThrSerLysSerAsnArgPheAlaPheAlaCysPro 189
      ::::: ||| ||| |||
QY 203 -----TCTGAGCTAGTCTGAAGGTGTCAGAGTTC 232
      ::::: ||| ||| |||
Db 190 ArgPheSerLysSerAspValCysCysSerProGlnAlaArgLeuSerLeuProGluGln 209
      ::::: ||| ||| |||
QY 233 AATGTCAGTCTCAAGATCAGGAGAAAGCTGGTCTCCAGATCTGCTTGAGCCTGTT 292
      ::::: ||| ||| |||
Db 210 SerLeuGlySerSerProSerProIleSerValThrAsnAspValTyrAlaLeuPhe 229
      ::::: ||| ||| |||
QY 293 AAAAATTCATCTTCTTTGGCCACTGTCAAAAGCAACTGAGTAGAGTCANATCAAGAAN 352
      ::::: ||| ||| |||
Db 230 AspSerSerAlaSer----- 234
      ::::: ||| ||| |||
QY 353 ACAGTGGAGTTACTCTGAACAAGAGAGATTGAACGGATCCACAGAGAAATAGCATTC 412
      ::::: ||| ||| |||
Db 235 -----ProLeuHisAlaGlyGluLeuSerSerLeuProGlyAlaValSerAla 250
      ::::: ||| ||| |||
QY 413 ATAAAGCG-----ACAAGTCTCTCCAAATGGACCTGCTGCTGAAGAACGGCA 466
      ::::: ||| ||| |||
Db 251 SerGluArgLeuLeuThrAlaProAlaGluIleGlyProSerAlaSerAlaCysLeu 270
      ::::: ||| ||| |||
QY 467 GGCAGAGCAGCTGGTTTTCCCTGGAGAAAGAGGAGCCAGCCATTGCTCCCATTAACA 526
      ::::: ||| ||| |||
Db 271 SerValSerCysGly-----ProGlyGluMet-SerProThrAlaAspThrArgHI 288
      ::::: ||| ||| |||
QY 527 TGTGCTCAGTGGCTGGAAGCAGCAACTCCCTGGAGCAGCAAAATTTCAACCTCCTCCA 586
      ::::: ||| ||| |||
Db 288 sasPalaGluGluArgGluArgArgAlaGluGluGluLys-----G1 303
      ::::: ||| ||| |||
QY 587 TAAGACAAGCAGCCAGTACAGACCTTTACTGACCCCTGTGMAAAGCCCTCTCTCCG 646
      ::::: ||| ||| |||
Db 303 uArgGluArgGln-----GluGluGluGluArgGluArgArg 315
      ::::: ||| ||| |||
QY 647 AGCCATGAGCTAGAGAGGCAAGATGCGACGAGCAGCTTCAGAGCGCTCGGCTCT 706
      ::::: ||| ||| |||
Db 315 gArgValGluGluGluGluArgGluArgGlnGluGluGluGluArgGluArgArg 335
      ::::: ||| ||| |||
QY 707 GCAGTCTACTATGANGCCAGGCTCGAAGAGAGAAATCNAAAGTTAAAGATATCA 766
      ::::: ||| ||| |||
Db 335 gValGluGluGluLysAlaArgGlnArgGluGluAspGlu-----ArgGluArg 351
      ::::: ||| ||| |||
QY 767 CAAGTCTGGAAGAAAGGCAAGAGCCCTAAAGAGTTTCAGCAGCTCGCGAA 826
      ::::: ||| ||| |||
Db 351 gArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArg 371
      ::::: ||| ||| |||
QY 827 GGTTAATCCAGCTCCGCTAGAGAACCAAGCAAGAAAGAGAAAGAGAGGAGGAGA-- 884
      ::::: ||| ||| |||
Db 371 gValGluGluGluLysAlaArgGlnArgGluGluGluGluGluArgGluArgArgVa 391
      ::::: ||| ||| |||
QY 885 -----AAGAAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 940
      ::::: ||| ||| |||
Db 391 lGluGluGluLysAlaArgGlnArgGluGluGluGluGluArgGluArgArgValG1 411
      ::::: ||| ||| |||
QY 941 AGAAGAGAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1000
      ::::: ||| ||| |||
Db 411 uGluGluLysAlaArgGlnArgGluGluGluGluGluGluGluArgGluArgValGlu 431
      ::::: ||| ||| |||
QY 1001 AAGAAGAACTAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1060
      ::::: ||| ||| |||
Db 431 uGlu-----LysAlaArgGlnArgGluGluGluGluGluGluGluArgGluArgValG1 449
      ::::: ||| ||| |||
QY 1061 AGAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120
      ::::: ||| ||| |||
```

```
Db 449 uGluGluLysAlaArgGlnArgGluGluGluGluGluArgGluArgValGluGlu 469
      ::::: ||| ||| |||
QY 1121 AAGAAAGAAAGAAATAAGACACNAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1163
      ::::: ||| ||| |||
Db 469 uLysGluArgGluArgGlnGluGluGluGluGluArgGluArgValGluGluLysG1 489
      ::::: ||| ||| |||
QY 1164 -AATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1222
      ::::: ||| ||| |||
Db 489 uArgGluArgGlnGluGluGluGluGluArgGluArgValGluGluLysGluArg 509
      ::::: ||| ||| |||
QY 1223 AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1282
      ::::: ||| ||| |||
Db 509 gGluArgGlnGluGluGluLysArgGluArgGluGluGluGluLysAlaArgG1 529
      ::::: ||| ||| |||
QY 1283 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298
      ::::: ||| ||| |||
Db 529 nArgGlnGluGluGlu 534
      ::::: ||| ||| |||
RESULT 12
AAU25510
ID AAU25510 standard; Protein; 611 AA.
XX
AC AAU25510;
XX
DT 17-DEC-2001 (first entry)
XX
DE T. gondii immunogenic protein POC-23.
XX
KW Immunogenic protein; oocyst; faeces; enteric apicomplexa oocyst;
KW Cryptosporidium oocyst; Toxoplasma oocyst; Giardia cyst; vaccine;
KW oocyte shedding.
XX
OS Toxoplasma gondii.
XX
PN US2001014447-A1.
XX
PD 16-AUG-2001.
XX
PF 18-DEC-1998; 98US-0216393.
XX
PR 19-DEC-1997; 97US-0994825.
XX
PA (MILH/) MILHAUSEN M J.
XX
PI Milhausen MJ;
XX
DR WPI: 2001-529100/58.
XX
DR N-PSDB; AAS42565.
XX
PT Detecting parasite oocysts or cysts in faeces, comprises eluting DNA
PT from sample into aqueous solution by heating, amplifying DNA with
PT primers specific for oocysts or cysts being detected, and detecting
PT amplification product.
XX
PS Example 2; Page 91-93; 188pp; English.
XX
CC The invention relates to detection of parasite oocysts or cysts in
CC a faeces sample comprising contacting the sample with a solid support,
CC drying and then washing the sample with an aqueous wash solution, adding
CC an aqueous elution solution and eluting DNA from the sample by heating
CC and amplifying by PCR oocyst/cyst-specific DNA and detecting the
CC amplification products. The method is useful for detecting parasite
CC oocysts e.g., enteric apicomplexa oocysts such as Cryptosporidium oocysts
CC or Toxoplasma oocysts, or for detecting parasite cysts e.g. Giardia
CC cysts. The method is also useful for developing vaccines to prevent
CC oocyte shedding in cats. The present sequence represents an
CC immunogenic protein from Toxoplasma gondii.
SQ Sequence 611 AA;
Alignment Scores: 3.05e-21 Length: 611
Pred. No.: 313.00 Matches: 100
Score:
```

Percent Similarity:	45.52%	Conservative:	103
Best Local Similarity:	22.42%	Mismatches:	154
Query Match:	14.63%	Indels:	89
DB:	22	Gaps:	11

US-09-502-945-3 (1-1298) x AAU25510 (1-611)

QY	107	GATGAGGGGACAATGATGGAGAGAAAGCATCAAAAGCTTCTGGAA-----	154
Db	130	AsnAspGlyAspArgGluGlyAspGluLysGluThrArgAspValGluAspGluGlyGlu149	
QY	155	-----GCAATCAGTCCCTTCATCGAAGAAGAAATAGCGGAAATTGGCTGAN199	
Db	150	ThrArgArgSerSerPheAlaGluGlnThrGlyAsnGluArgThrGluMetArgThr169	
QY	200	AGG-----	202
Db	170	ArgHisGlyGlyAspGluGlyTrpThrSerLysSerAsnArgPheAlaPheAlaCysPro189	
QY	203	-----TCTGAGCCTAGTCTGAGGTGCAGAGTTCAGAGTTC232	
Db	190	ArgPheSerLysSerAspValCysCysSerProGlnAlaArgLeuSerLeuProGluGln209	
QY	233	AATGTCAGTCTTGAAGGATCAGGAGAAAGCTGGTCTTCAGATGTCGTGTGAGCCTGTT292	
Db	210	SerLeuGlySerSerProSerSerProIleSerValThrAsnAspValTyAlaLeuPhe229	
QY	293	AAAACTTCATCTTCTTTGGCCACTGTGAAAAGCAACTGAGTAGCTCANATCAAGAAN352	
Db	230	AspSerSerAlaSer-----	234
QY	353	ACAGTGGAGTTACCTCTGACAAAGAGAGATTGAACGGATCCACAGAAATAGCATTC412	
Db	235	-----ProLeuHisAlaGlyLeuLeuSerSerLeuProGlyAlaValSerAla250	
QY	413	ATAAAACGC-----ACAAGTCCTCTCCAATGGGACCTGCTGCTCTGAAGACCGCA466	
Db	251	SerGluArgLeuLeuThrAlaProAlaGluIleGlyProSerAlaSerSerAlaCysLeu270	
QY	467	GGCAGACAGCTGTTTTTCCCTTGGAGAGAGAGCCAGCCATTCCTCCCATGGAACA526	
Db	271	SerValSerCysGly-----ProGlyGluMet-SerProThrAlaAspThrArgHis288	
QY	527	TGTGCTCAGTGGCTGGAAGCAGAACTCCCTCGGAGCAGGAATTTTCAACCTCCTCCA586	
Db	288	sAspAlaGluArgGluArgArgAlaGluGluGluLys-----G1303	
QY	587	TAAGAACAAAGCAGCAGTGCAGACCCCTTTACTGACCCCTGTGGAAAAGCCCTCTCTCG646	
Db	303	uArgGluArgGln-----GluGluGluGluArgGluArgGluArg315	
QY	647	AGCCATGAGCCTTAGAAGCAAAAGATGCGACGACGAGCTTCAGAGGCTCGGGCTCT706	
Db	315	gArgValGluGluLysGluArgGluArgGlnGluGluGluArgGluArgGluArgGluArg335	
QY	707	GCAGTCTCTATTANGCCCAAGCTCGAAGAGAGAGAAAATCAAAAGTTAAAGTATCA766	
Db	335	gValGluGluLysAlaArgGlnArgGluGluAspGlu-----ArgGluArg351	
QY	767	CAAAAGTCGTGAAGAAAGGAGCCCAAGAAAGCCCTTAAAGAGTTTGACAGCTCGCGAA826	
Db	351	gArgArgValGluGluGluLysAlaArgGlnArgGluGluGluArgGluArgGluArgGluArg371	
QY	827	GGTTAATCCAGCTCGCGCCTAGCAACCAAGAAAAGAGGAAAGAGGAGGAGGAGA--884	
Db	371	gValGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgGluArgGluArgGluArg391	
QY	885	-----AGAAGAAGACAAAGAGAAAGAAAGAAAGGAGGAGGAGGAGGAGGAGGAG940	
Db	391	lGluGluGluLysAlaArgGlnArgGluGluGluArgGluArgGluArgGluArgGluArg411	
QY	941	AGAAGAGGAAAGCAGAGGAGGAGAAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGAA1000	

[illegible]

CC at ftp.wipo.int/pub/published_pct_sequences.

XX	Sequence	2274 AA;
SQ		

Alignment Scores:

Pred. No.:	5.77e-21	Length:	2274
Score:	312.00	Matches:	128
Percent Similarity:	47.31%	Conservative:	83
Best Local Similarity:	28.70%	Mismatches:	136
Query Match:	14.58%	Indels:	99
DB:	22	Gaps:	16

US-09-502-945-3 (1-1298) x ABB58657 (1-2274)

Qy	101	AGTGAAGATGAGGGGACAAATGATGAGGAGAGAAAGCATCNAAGCTTCTGGAAGCAATC	160
Db	1203	SerGluAspGluAlaThr-----ThrThrIleThrLeuAlaLys	1215
Qy	161	AGTTCCCTTGATGGAAGAAGATAGCGGAAATGGCTGANAGTCTGAGGCTAGTCTCAAG	220
Db	1216	LeuAlaLysLysAlaLysAsnAsnLysGln-----LysThrSerGlyLys	1230
Qy	221	GTGTCAAGATTCAATGCTAGTCTGAAGGATCAGGAGAAAGCTGGTCTTCTGCAGATCTG	280
Db	1231	ValSerGluLysLysProGluGlnThrAsnLeuAlaGluGluLysVal-----	1246
Qy	281	CTTGAGCCTGTAAACCTCATCTCTTTGGCCACTGTGTAAGGAGCAAAAGCAACTGAGTAGATC	340
Db	1247	-----AlaValGluLysGluGluAsnGluGlu	1255
Qy	341	ANATCAAGAANACAGTGGAGTTACCTCTGACAAAGAA-----GAGATTGAACGG	391
Db	1256	ArgAlaValProMetValLysThrIleAsnLysGluGlnAspSerAlaGluSer	1275
Qy	392	ATCCACAGAGAAATGACATTAATAAACCCACAGCTCTCCCAATGGGACCTCTCGT	451
Db	1276	AspHisAlaAsp-SerLeuLeuAlaAsnLysSerSerIleAla-----AlaValMe	1292
Qy	452	CCTGAAGAAGCGCAGCAGCAGCAGCTGGTGTTCCTCGGAGAGAGAGCCAGCCAT	511
Db	1292	tValSerSerAlaSerAlaGlnGlyLeuSerLeuHisValGlu-----	1306
Qy	512	TGCTCCCATTCACATGCTGCTACGTGGCTGGGAAGCAAGCAACTCCCTGGAGCAGGAAT	571
Db	1307	-----MetSerAlaAlaAspAlaGluGlnGlyGluAspGluGluI	1320
Qy	572	TTTCAAACCTC-----CTCCATAAGACAAGACGACC	601
Db	1320	eGluGlyLeuAspGluGluProProLysThrMetSerLysAspAsnLysLysLysGlnL	1340
Qy	602	AGTCAGACACCTTTACTGACCCTGTG-----GAAAAGCCCTCTCTCCG	646
Db	1340	sProGlyAspAlaValAlaThrMetThrIleAspLysGluLysAlaLysGluL	1360
Qy	647	AGCCATGAGCTGAAGACCAAGATGCGACGACGAGCAAGCTTCAG-----	692
Db	1360	sGluLeuLysLeuLysGlu--LysGluArgGluAlaLysLeuGlnLysGluLysG	1379
Qy	693	-----AGGGCTCGGGCTCT	706
Db	1379	uGluLysLeuLysLeuLysGluArgGluGluSerLeuArgMetGluArgGluGluLysLe	1399
Qy	707	CGAGCTCTACTATGAGCCAGGCTCGAGAGAGAGAAATCNAAGATTAAAGATATCA	766
Db	1399	uLysGluGluLysLysGluLysGluArgGluGluLysLeu-----LysGluG	1416
Qy	767	CAAGTCTGTAAGAAGGAAGCGCAAGAGCCCTAAAAGAGTTGTGACGCTCGCGAA	826
Db	1416	uLysIleLysGluLysGlnArgGluGluLys--LeuLysGlu--GluLysLeuLysG	1434
Qy	827	GETTAATCCAGCTCCGCACTAGAGAAGCAAGAAAGAGGAAGA-----AGGAG	877
Db	1434	uLysGluArgGluGluArgMetLysGluLysGluArgGluLysAlaLysGluLysG	1454

Qy	878	GAGGAGAAGAAGAACAACGAGGAAGAAGAAGAAGGGAGCAAGCAGAGAAGAAGA	937
Dc	1454	nArGluLuLysLeuArgGlUGluLysIleLysGluLysGluArgGlu-GluLysLeuL	1474
Qy	938	AGGAGAAGAGAAAGGAAGAGGAGCAAAAAGAAAGAGAGAA-----AAGG	985
Dc	1474	ysGluLysLeuArgGluGluLysIleLysGluLysGluLysGluLylsLeuArgLysG	1494
Qy	986	AAAAAGAGGAGAAGAAGAACTTAAGAAGAAGGAGAGGAAAGATAAGAAGAAAG	1045
Dc	1494	luArgGluGluLysMetArgGlu--LysGluArgGluGluLysIleLysGluLysGluAr	1513
Qy	1046	AAAGAAAAATNNAAGAACAGAAAGAGGAGGAAGGAAGGAAGCAAGGAGCAACTN	1105
Dc	1513	gvalGluLysIleLysGluLysGluArgGluGluLysLeuLysGluLysGluLyls	1533
Qy	1106	GAAGAAAGAGGAGCAAGAAAGAAAGAACTAAGCAACNAGAAAGAAAGGAGCAAG	1165
Dc	1533	sLeuLysGluLysGluLysLeuLysLysLysGluLysGluLysGluArgGluGlu	1553
Qy	1166	TAAG---AAGAGGAAGAAAGAAAGAAAGAAAGCAAGGAGGAAGGAGGAAAG	1222
Dc	1553	uLysLeuLysGluLysGluArgGlnGluLysLeuLysGluLysGluArgGluGlu	1573
Qy	1223	AAAAAGAGGAAGTAGAAGCGGAGGAAGAAAGAAAGAAAGTATAAGAGGAAAG	1288
Dc	1573	uLysArgGluThrGluGluArgGlnArgGluLysGlu-ArgGluGluLysLeuLys	1593
Qy	1283	AAGAAAGA 1290	
Dc	1593	ysGluArg 1595	
 RESULT 14 ABB62436 ID ABB62436 standard; Protein; 771 AA.			
XX	AC	ABB62436;	
XX	DE	26-MAR-2002 (first entry)	
XX	KW	Drosophila melanogaster polypeptide SEQ ID NO 14100.	
XX	KW	Drosophila; developmental biology; cell signalling; insecticide;	
XX	OS	Drosophila melanogaster.	
XX	PN	WO200171042-A2.	
XX	PD	27-SEP-2001.	
XX	PF	23-MAR-2001; 2001WO-US09231.	
XX	PR	23-MAR-2000; 2000US-191637P.	
XX	PA	(PEKE) PE CORP NY.	
XX	PI	Venter JC, Adams M, Li PWD, Myers EW;	
XX	DR	WPI; 2001-656860/75.	
XX	N-PSDB	ABL06539.	
PT	PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -	
PS	PS	Disclosure; SEQ ID NO 14100; 2lpp + Sequence Listing; English.	
CC	CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from drosophila. The invention is useful in developmental biology and in elucidating cell signalling and	

PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX
123456789

PS Claim 20; SEQ ID No 56902; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 405 AA;

Alignment Scores:

Pred. No.:	6,47e-21	Length:	405
Score:	309.00	Matches:	83
Percent Similarity:	59.75%	Conservative:	61
Best Local Similarity:	34.44%	Mismatches:	56
Query Match:	14.44%	Indels:	41
DB:	22	Gaps:	7

US-09-502-945-3 (1-1298) x ABG26543 (1-405)

[illegible]

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:34:32 ; Search time 10.0556 Seconds
(without alignments)
7595.986 Million cell updates/sec

Title: US-09-502-945-3
Perfect score: 2140
Sequence: 1 ggcgtggaactgactgcga.....gaagaagaagaagagaaaaa 1298

Scoring table: BLOSUM62

Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p_model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09502945/runat_14032003_101059_19145/app_query.fasta_1.10979
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPL=0
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_80 -runat_14032003_101059_19145 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued_Patents_AA.*
- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	401	18.7	434	1	US-08-097-830B-3
2	401	18.7	434	2	US-08-456-112B-3
3	298	13.9	1898	1	US-08-056-200-94
4	298	13.9	1898	2	US-08-800-644-94
5	265	12.4	100	2	US-08-460-890A-62
6	265	12.4	100	3	US-08-167-641C-62
7	265	12.4	100	4	US-08-460-971A-62
8	265	12.4	100	4	US-08-462-040-62
9	261	12.2	1507	3	US-08-929-329-5
10	234.5	11.0	432	2	US-08-933-750C-47
11	234.5	11.0	432	4	US-09-234-613-47
12	211	9.9	558	1	US-08-285-440-6

13	211	9.9	558	1	US-08-630-349-6
14	209	9.8	257	4	US-09-461-697-188
15	209	9.8	272	4	US-09-461-697-186
16	205	9.6	231	4	US-09-461-697-194
17	205	9.6	232	4	US-09-461-697-192
18	205	9.6	238	4	US-09-461-697-190
19	201	9.4	532	1	US-08-285-440-5
20	201	9.4	532	1	US-08-630-349-5
21	200.5	9.4	300	4	US-09-395-689-1
22	200.5	9.4	765	2	US-08-663-112-2
23	198.5	9.3	1588	5	PCT-US93-07261-11
24	198.5	9.3	1663	5	PCT-US93-07261-16
25	189	8.8	683	6	5210183-3
26	188.5	8.8	290	2	US-08-903-801-1
27	188.5	8.8	290	2	US-09-295-055-1
28	188.5	8.8	614	5	PCT-US95-03236-21
29	185.5	8.7	288	2	US-08-961-858-6
30	185.5	8.7	288	3	US-09-089-593-6
31	185.5	8.7	288	3	US-08-950-925-4
32	180.5	8.4	1360	4	US-09-393-569-2
33	179.5	8.4	118	4	US-09-134-001C-2856
34	179.5	8.4	316	4	US-08-098-327E-31
35	179.5	8.4	316	4	US-08-462-625-31
36	178	8.3	631	4	US-08-847-065-25
37	177	8.3	234	2	US-08-903-801-3
38	177	8.3	234	4	US-09-295-055-3
39	177	8.3	564	4	US-09-308-022-6
40	176.5	8.2	816	2	US-08-533-306A-6
41	176.5	8.2	816	2	US-08-742-923A-6
42	176.5	8.2	885	2	US-08-533-306A-4
43	176.5	8.2	885	2	US-08-742-923A-4
44	175	8.2	514	2	US-08-960-022-14
45	174	8.1	564	2	US-08-216-894-2

ALIGNMENTS

RESULT 1
US-08-097-830E-3
; Sequence 3, Application US/08097830E
; Patent No. 5652211
; GENERAL INFORMATION:
; APPLICANT: Porro, Massimo
; TITLE OF INVENTION: Peptides For Neutralizing The
; TITLE OF INVENTION: Toxicity of Lipid A
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hedman, Gibson & Costigan, P.C.
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/097,830E
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Costigan, James V.
; REGISTRATION NUMBER: 25,669
; REFERENCE/DOCKET NUMBER: 576-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 302-8989
; TELEFAX: (212) 302-8998
; INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 434 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-097-830E-3

Alignment Scores:
Pred. No.: 8.34e-32 Length: 434
Score: 401.00 Matches: 79
Percent Similarity: 73.33% Conservative: 86
Best Local Similarity: 35.11% Mismatches: 56
Query Match: 18.74% Indels: 4
DB: 0 Gaps: 0

US-09-502-945-3 (1-1298) x US-08-097-830E-3 (1-434)

QY 630 GAAAGGCTCTCCGAGCCATGAGCTAGAGAGGCAAGATCGCAGCAGCAGCTT 689
Db 1 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 20
QY 690 CAGAGGCTCGGCTCTGCGAGTCTACTATGANGCCAGGCTCGAAGAGAGAAATC 749
Db 21 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
QY 750 NAAAGTTAAAGATCATCAAGTCTGNAGAGAGAGAGGCGCAAGAGCCCTAAAGAG 809
Db 41 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 60
QY 810 TTTGAGCAGCTGCGGAGGTTAATCCAGCTCGCCAGCTAGAGAGAGAGAGAGAA 869
Db 61 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
QY 870 AGAGGAGGAGGAG 929
Db 81 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100
QY 930 AGAAAGAGGAA 989
Db 100 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 120
QY 990 GAAGGAA 1049
Db 120 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 139
QY 1050 AAAAAAGTAAAG 1109
Db 140 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 159
QY 1110 AAAGAGGAA 1169
Db 160 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 179
QY 1170 AAGAGGAA 1229
Db 180 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 199
QY 1230 GAAGAA 1289
Db 199 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 219
QY 1290 AAGGAAA 1296
Db 219 sLysLys 221

RESULT 2
US-08-456-112B-3
Sequence 3, Application US/08456112B
Patent No. 5834430
GENERAL INFORMATION:
APPLICANT: Porro, Massimo
TITLE OF INVENTION: POTENTIATION OF ANTIBIOTICS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:

ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: LEADING EDGE 486
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,112B
FILING DATE: May 31, 1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8998
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 434 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-08-456-112B-3

Alignment Scores:
Pred. No.: 8.34e-32 Length: 434
Score: 401.00 Matches: 79
Percent Similarity: 73.33% Conservative: 86
Best Local Similarity: 35.11% Mismatches: 56
Query Match: 18.74% Indels: 4
DB: 0 Gaps: 0

US-09-502-945-3 (1-1298) x US-08-456-112B-3 (1-434)

QY 630 GAAAGGCTCTCTCGAGCCATGAGCTAGAGAGGCAAGATCGCAGCAGCAGCTT 689
Db 1 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 20
QY 690 CAGAGGCTCGGCTCTGCGAGTCTACTATGANGCCAGGCTCGAAGAGAGAAATC 749
Db 21 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
QY 750 NAAAGTTAAAGATCATCAAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 809
Db 41 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 60
QY 810 TTTGAGCAGCTCGGAGGTTAATCCAGCTCGCCAGCTAGAGAGAGAGAGAGAA 869
Db 61 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
QY 870 AGAGGAGGAGGAG 929
Db 81 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100
QY 930 AGAAAGAGGAA 989
Db 100 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 120
QY 990 GAAGGAA 1049
Db 120 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 139
QY 1050 AAAAAAGTAAAG 1109
Db 140 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 159


```

QY 1238 AAGTACGAACCCGAGAACAAGAAAGAAAGATTTATAAGAGAGAACAGAAAGAAAAGG 1293
    || ||| |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
Db 456 ySHisGIuGlncIuArgArgGluGlnArgLeuLysArgGluGlnGluArgArg 474

RESULT 5
US-08-460-890A-62
; Sequence 62, Application US/08460890A
; Patent No. 5994109
; GENERAL INFORMATION:
; APPLICANT: Woo, Savio L.C.
; APPLICANT: Smith, Louis C.
; APPLICANT: Cristiano, Richard J.
; APPLICANT: Gottchalk, Stephen
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND
; TITLE OF INVENTION: METHODS OF USE
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,890A
; FILING DATE: June 5, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167,641
; FILING DATE: December 14, 1993
; APPLICATION NUMBER: 07/855,389
; FILING DATE: March 20, 1992
; APPLICATION NUMBER: PCT/US93/02725
; FILING DATE: March 19, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 212/066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: "Lys" in positions 3 to 100 may be present
; OTHER INFORMATION: or absent.
US-08-460-890A-62

Alignment Scores:
Pred. No.: 1,76e-18 Length: 100
Score: 265.00 Matches: 48
Percent Similarity: 89.32% Conservative: 44
Best Local Similarity: 46.60% Mismatches: 8
Query Match: 12.38% Indels: 3
DB: 2 Gaps: 0

US-09-502-945-3 (1-1298) x US-08-460-890A-62 (1-100)

QY 934 AAGNAGAGAGAGAGGAAAAAGAGACAGAGGAGAAAGAGAAAGAGAAAGAAAAAGAG 993
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```



```

Db 183 AlaAlaGlnGluLysGlnMetGluValCys----- 192
QY 545 GGCAAGAACTCCCTGGAGCAGGAAATTTCAACCTCTCCATAGAACAGCAGCCAGT 604
Db 192 ----- 192
QY 605 GACAGACCCCTTACTGACCCCTGTGGAAGAGCCCTCTCCGAGCCATGAGCCCTAGAAGA 664
Db 193 -----GluValCysGlyAlaPheLeuIleValGly----- 202
QY 665 GGCAAGAACTGGAGCAGCAGCAGCTTCAGAGGCTCGGCTCTGCAGCTCTACTATGANGC 724
Db 203 -----AspAlaGlnSerArgValAspHis-----Leu 212
QY 725 CAAGGCTCGAAGACAGAGAAATCNAAGTTAAA-----GTATCACAAGTCGGA 778
Db 213 MetGlyLysGlnHisMetGlyTyrAlaLysIleLysAlaThrValGluLeuLysGlu 232
QY 779 GAA---AGGAAAGGCCAAGAGCCCTAAAGAGTTTGAGCAGCTCGGAGGTTTATCC 835
Db 233 LysLeuArgLysArgThrGluGluProasp----- 242
QY 836 AGTGGCCGCTAGAGAAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 895
Db 243 -----ArgAspGluArgLeuLysLysGluLysGlnGluArgGluGluArg 257
QY 896 CAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 955
Db 258 GluLysGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGlu 277
QY 956 AGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1015
Db 278 GluArgGluLysGluArgAlaArgAspArgGluArgLysArgSer-ArgSerArgSe 297
QY 1016 GAAG-----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1057
Db 297 rArgHisSerSerArgThrSerAspArgArgCysSerArgSerArgAspHisLysArgSe 317
QY 1058 NAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
Db 317 rArgSerArgGluArgArgThrArgSerArgAspArgArgSerArgSerHisAs 337
QY 1118 G-----GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1171
Db 337 pArgSerGluArgLysHisArgSerArgSerArgSerArgSerArgSerArgSerLys 357
QY 1172 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1231
Db 357 pArgLysSerTyrLysHisArgSerLysSerArgAspArgGluGlnAspArgLysSerLy 377
QY 1232 AGAAGAACTAGAACGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1285
Db 377 sGluLysGluLysArgGlySerAspAspLysLysSerValLysSerGlySerArgG1 397
QY 1286 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298
Db 397 uLysGlnSerGlu 401

RESULT 11
US-09-234-613-47
; Sequence 47, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMNOT04
; CLONE: 2926777
; US-09-234-613-47

Alignment Scores:
Pred. No.: 3,39e-15 Length: 432
Score: 234.50 Matches: 91
Percent Similarity: 41.23% Conservative: 76
Best Local Similarity: 22.47% Mismatches: 128
Query Match: 10.96% Indels: 110
DB: 4 Gaps: 15

US-09-502-945-3 (1-1298) x US-09-234-613-47 (1-432)
QY 149 CTGGAAGCAATCAGTCCCTTGATGGAAAGATAGCGGAAATTGGCTGANNAGTCTGAG 208
Db 86 LeuArgTyrLeuGlnSerLeuLeuAlaGluValGluArgGlyLeu---ArgArgGlyHis 104
QY 209 GCTAGTCTGAAGGTGTCAGAGTTCAATGTCAGTCTGAGGATCAGGA----- 256
Db 105 AlaArgLeuAlaLeuSerGlnAsnGlnGlnSerSerGlyAlaAlaGlyProThrGlyLys 124
QY 257 -----GAAAGCTG---GTCTTTGCGATCTCTGAGCCTGTTAAACATTCATCTTCT 307
Db 125 AsnGluGluLysIleGlnValLeuThrAspLysIleAsp----- 137
QY 308 TTGGCCACTGTGAAAGCAACTGAGTAGCTAGCTCANNATCAAGAAACACATGGAGTTACCT 367
Db 138 -----ValLeuLeuGlnGlnIleGluLeuLysSerGlyLysValGluAla 155
QY 368 CTGAACAAAAGAGAG---ATTGAACGATCCACAGAGATAGCATTCATAAAGCCACA 424
Db 156 GlnGlyMetMetLysLeuValGluGlnLysGluGluArgGluLeuArgSerThr 175
QY 425 AGTCCTCTCCAAATGGGACCCTGCTGCTGAGAACCGGAGGACAGCAGCTGTTT 484
Db 176 Thr-----SerThrIleGluSerPhe 182
QY 485 TCCCTCGGAGAAAGAGAGGAGCCCATTTGCTCCCATTTGAACATGTGCTCAGTGGTGAA 544

```

183 AlaAlaGlnGluLysGlnMetClnuValCys----- 192

545 GCACAAGAACTCCCTCGGAGCAGGAATAATTTCAACCTCTCCATAAGAACAAGCAGCCAGT 604

192 ----- 192

605 GACACACCTTTACTGACCCCTGTGAAAGGCGCTCTCTCCGAGCCATGAGCCTAGAGA 664

193 ----- 202

665 GCACAAGATGCGACGAGCAGCAGCTTCAGAAGGCTCGGCTCTGCAGTCTCTACTATGANGC 724

203 -----AspAlaGlnSerArgValAspHis-----Leu 212

725 CAAGCTCGAAGCAGAGAGAAATCNAAGTTAAA-----GTATCACAAGTCGTAA 778

213 MetGlyLysGlnHisMetGlyTyrAlaLysIleLysAlaThrValGluGluLeuLysGlu 232

779 GAA---AGGAAGGCGCAAGAAAGCCCTAAAGAGTTTGAGCAGCTGCGGAGGTTAATCC 835

233 LysLeuArgLysArgThrGluLysProasp----- 242

836 AGCTGCCGCCTAGAGAAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 895

243 -----ArgAspGluArgLeuLysLysGluLysGlnGluArgGluGluArg 257

896 CAAGCAGAGAGAAAG 955

258 GluLysGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluGlu 277

956 AGAAGCAG 1015

278 GluArgGluLysGluArgAlaArgAspArgGluArgLysArgLysArgSerArgSerArgSe 297

1016 GAAG-----GAG 1057

297 rArgHisSerSerArgThrSerAspArgArgCysSerArgSerArgSerArgAspHisLysArgSe 317

1058 NAAAG 1117

317 rArgSerArgGluArgArgArgThrArgSerArgSerArgSerArgSerArgSerArgHis 337

1118 G-----GAAAG 1171

337 pArgSerGluArgLysHisArgSerArgSerArgSerArgSerArgSerArgSerArgSer 357

1172 GAG 1231

357 pArgLysSerTyrLysHisArgSerLysSerArgAspArgGluGlnAspArgLysSerLy 377

1232 AGAAG 1285

377 sGluLysGluLysArgGlySerAspAspLysLysSerSerValLysSerGlySerArgG1 397

1286 AAAGAAGAGAGAG 1298

397 uLysGlnSerGlu 401

RESULT 12

US-08-285-440-6

; Sequence 6, Application US/08285440

; Patent No. 5532337

; GENERAL INFORMATION:

; APPLICANT: Ken'ichiro HAYASHI et al.

; TITLE OF INVENTION: POLYPEPTIDE

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wenderoth, Lind & Ponack

; STREET: 805 Fifteenth Street, N.W., #700

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 500 kb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: DisplayWrite

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/285,440

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/858,947

FILING DATE: March 27, 1992

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-8850

TELEFAX:

TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 558 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL:

ANTI-SENSE:

FRAGMENT TYPE:

ORIGINAL SOURCE:

ORGANISM:

STRAIN:

INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE:

HAPLOTYPE:

TISSUE TYPE:

CELL TYPE:

CELL LINE:

ORGANELLE:

IMMEDIATE SOURCE:

LIBRARY:

CLONE:

POSITION IN GENOME:

CHROMOSOME/SEGMENT:

MAP POSITION:

UNITS:

FEATURE:

NAME/KEY:

LOCATION:

IDENTIFICATION METHOD:

OTHER INFORMATION:

PUBLICATION INFORMATION:

AUTHORS:

TITLE:

JOURNAL:

VOLUME:

ISSUE:

PAGES:

DATE:

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:

US-08-285-440-6

Alignment Scores:

Pred. No.: 8.25e-13

Score: 211.00

Percent Similarity: 40.17%

Best Local Similarity: 25.14%

Query Match: 9.86%

Length: 558

Matches: 87

Conservative: 52

Mismatches: 88

Indels: 119

DB:	1	Gaps:	12
US-09-502-945-3 (1-1298) x US-08-285-440-6 (1-558)			
Qy	579	CTCCTCCATAGACACAGCAGCAGCTACAGACCCCTTTACTGACCCCTGTGGAAAAGGCC	638
Db	44	LeuArgGlnLysGlnGluGluGluSerLeuGlyGlnValThrAspGlnValGluValAsn	63
Qy	639	TCTCTCGAGCCTAGACCTTAGAAGAGGCAAGATCGCAGCAGCAGCTTCAG	692
Db	64	AlaGlnAsnSerValProAspGluGluAlaLysThrThrThrAsnThrGlnValGlu	83
Qy	693	-----AGGCTCGGGCTCTGCAGTCTCTACTATGANGCAAGCTCGAGAGAGAAG	743
Db	84	GlyAspAspGluAlaAlaPheLeuGluArgLeuAlaArgArgGluArgGlnLys	103
Qy	744	AAATCTAAAGTTAAAGTATCACAAGTCGTCAAGAAAGGAAGCCCAAGAAAGCCCTA	803
Db	104	ArgLeuGln	112
Qy	804	AAAGAGTTTGAG	830
Db	113	LysGluPheAspProThrIleThrAspAlaSerLeuSerLeuProSerArgArgMetGln	132
Qy	831	AATCCAGCTCCGCACTAGAAGAACGAGAAAGAGAAAGAGGAGGAGGAGAAG	887
Db	133	AsnAspThrAlaGluAsnGluThrThrGluLysGluGluLysSerGluSerArgGlnGlu	152
Qy	888	-----AAGAACAACAAGGAGAG	905
Db	153	ArgTyrGluIleGluGluThrGluThrValThrLysSerTyrGlnLysAsnAspTrpArg	172
Qy	906	AAGAAGAAGAGGAGAGGAGAGAGAAAGAGAGAGAGAGAGAAAGAGAGAA	959
Db	173	AspAlaGluGluAsnLysLysGluAspLysGluLysGluGluGluGluLysPro	192
Qy	960	-----CGAGAAACAAGAGAGAGAGAAAGAGAAAGAGAGAGAGAGAGAA	1004
Db	193	LysArgGlySerIleGlyGluAsnGlnGlyGluLysGlyThrLysValGlnAlaLys	212
Qy	1005	AGAACT	1037
Db	213	ArgGluLysLeuGlnGluAspLysProThrPheLysLysLysGluGluLysAspGluLys	232
Qy	1038	GAAGAAGAAGAAAGAAAGTAAAGAGAA	1079
Db	233	IleLysLysAspLysGluProLysGluGluValLysSerPheMetAspArgLysGly	252
Qy	1079	-----	1079
Db	253	PheThrGluValLysSerGlnAsnGlyGluPheMetThrHisLysLeuLysHisThrGlu	272
Qy	1080	-----AAGGAAGAAGAGGA	1094
Db	273	AsnThrPheSerArgProGlyGlyArgAlaSerValAspThrLysGluAlaGluGlyAla	292
Qy	1095	-----AGAACTNAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1133
Db	293	ProGlnValGluAlaGlyLysArgLeuGluGluLeuArgArgArgGlyGluThrGlu	312
Qy	1134	AATAGAACAACAG	1175
Db	313	SerGluGluPheGluLysLeuLysGlnLysGlnGluAlaAlaLeuGluLeuGluGlu	332
Qy	1176	AAGAAGAAAAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1235
Db	333	LeuLysLysArgGluGluArgArgLysValLeuGluGluGluGluGluGluGluGly	352
Qy	1236	GAAGTAGAAGCGGAGAGAG	1277
Db	352	SGlnGluAlaAspArgLysLeuArgGluGluGluLysArgArgLeuLysGluGlu	372
Qy	1278	AGAGAAGAAAGAGAGG	1293

```

Db      372 uileGluArgArg 377
      I      III      IIIII
RESULT 13
US-08-630-349-6
; Sequence 6, Application US/08630349
; Patent No. 5739008
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,349
; FILING DATE: April 10, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/285,440
; FILING DATE: August 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek Jr.
; REGISTRATION NUMBER: 33,367
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 558 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-630-349-6

Alignment Scores:
Pred. No.:      8,25e-13      Length:      558
Score:          211.00      Matches:      87
Percent Similarity: 40.17%      Conservative: 52
Best Local Similarity: 25.14%      Mismatches:   88
Query Match:      9.86%      Indels:      119
DB:              1          Gaps:       12

US-09-502-945-3 (1-1298) x US-08-630-349-6 (1-558)
Qy 579 CTCCTCCATAGAACACGACCGCAGTCGACAGACCTTTACTGCCCTGTGGAAGGCC 638
      III      III      ::::      :::      IIIII
Db 44 LeuArgGlnLysGlnGluGluGluSerLeuGlycInValThrAspGlnValGluValAsn 63
      :::      :::      :::      :::      :::
Qy 639 TCTCTCCGAGCCATGACCCCTAGACGAGCGCAAGATGCGACGACGACGAGCTTCAG----- 622
      :::      :::      :::      :::      :::      :::      :::      :::
Db 64 AlaGlnAsnSerValProAspGluGluAlaLysThrThrThrAsnThrGlnValGlu 83
      :::      :::      :::      :::      :::      :::      :::      :::
Qy 693 -----AGGGCTCGGGCTCTGCAGCTCCTACTATGANGCCCAAGGCTCGAAGAGAGAAG 743
      :::      :::      :::      :::      :::      :::      :::      :::
Db 84 GlyAspAspGluAlaAlaPheLeuGluArgLeuAlaArgArgGluGluArgGlnLys 103
      :::      :::      :::      :::      :::      :::      :::      :::
Qy 744 AAAATCNAAGTTAAAAAGTATCACAAAGTCGTGAAGAAAGGAAGGCCCAAGAGCCCTA 803
      :::      :::      :::      :::      :::      :::      :::      :::
Db 104 ArgLeuGln-----GluAlaLeuGluArgGln 112
      :::      :::      :::      :::      :::      :::      :::      :::
Qy 804 AAGAGGTTTGAG-----CAGCTGCGGAAGGTT 830
      :::      :::      :::      :::      :::      :::      :::      :::

```


RESULT 15

US-09-461-697-186
; Sequence 186, Application US/09461697
; Patent No. 6277374
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: LO, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 186
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-186

Alignment Scores:

Pred. No.:	9.93e-13	Length:	272
Score:	209.00	Matches:	77
Percent Similarity:	48.47%	Conservative:	50
Best Local Similarity:	29.39%	Mismatches:	71
Query Match:	9.77%	Indels:	65
DB:	4	Gaps:	11

US-09-502-945-3 (1-1298) x US-09-461-697-186 (1-272)

QY	600	CCAGTGACGACCCCTTTACTGACCCCTGGAAGAGGCTCTCCGAGCCATGAGCCTA	659
DB	19	ProValThr---ProGluValLysProLysArgThrSerSerArgLysMetLysThr	37
QY	660	GAAGAGGCAAGATCGCGACGA-----GCAGAGCTTCAG	692
DB	38	LysSerAspMetMetGluGluAsnIleAspThrSerAlaGlnAlaValAlaGluThrLys	57
QY	693	AGGCTCGGGCTCTGCAGCTCTACTATGACCAGGCTCGAAGA-----	737
DB	58	GlnGluAlaValAlaGluGluAspTyrAsnGluAsnAlaLysAsnGlyGluAlaLysIle	77
QY	738	-----GAGAAGAAATCNAAGTTAAAGTAAAGTATCACAAAGTC-----	773
DB	78	ThrGluAlaProAlaSerGluLysGluIleValGluValLysGluGluAsnIleGluAsp	97
QY	774	---GTGAAGAAAGGAAGGCCCAAGAA-----GCCCTAAAGAGTTTGAG	815
DB	98	AlaThrGluLysGlyGlyGluLysGluAlaValAlaAlaGluValLysAsnGluGlu	117
QY	816	CAGCTCGGAAGCTTAATCCAGCTGCCCGCTAGAACGA-----AGAAAAGAG	866
DB	118	GluAspGlnLysGluAspGluGluAspGlnAsnGluGluLysGlyGluAlaGlyLysGlu	137
QY	867	GAAGAAGGAGGAGGAAGAAGAACAAGGAG-----	902
DB	138	AspLysAspGluLysGlyGluGluAspGlyLysGluAspLysAsnGlyAsnGluLysGly	157
QY	903	-----AAGAAGAAAGAGGAGGAGAGG-----	926
DB	158	GluAspAlaLysGluLysGluAspGlyLysGlyGluAspGlyLysGlyAsnGlyGlu	177
QY	927	---AGAGAAAGAGGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	983
DB	178	AspGlyLysGlyGlyGluAspGluGluGluGluGluGluGluGluGluGluGluGlu	197
QY	984	GGAAGAGGAG	1043

DB	198	GlyLysGluAsnGluAsp-----GlyLysGluLysGlyAspLysLysGluGly	213
QY	1044	AGAAAGAAAAAGTAAAG	1099
DB	214	LysAspValLysValLysGluAspGluLysGluArgGluAspGlyLysGluAspGluGly	233
QY	1100	TNAGAAGAAG	1159
DB	234	GlyAsnGluGluAlaGlyLysGluLysGluAsp-----LeuLysGluGlu	249
QY	1160	AAAGAATAAG	1219
DB	250	GluGluGlyLysGluGluAspGluLysGlu-----AspAspGlyLysGluGluProG	269
QY	1220	AA 1221	
DB	269	ln 269	

Search completed: March 21, 2003, 13:08:07
Job time : 21.0536 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 13:00:02 ; Search time 13.8984 Seconds
(without alignments)
9985.861 Million cell updates/sec

Title: US-09-502-945-3
Perfect score: 2140
Sequence: 1 ggcgtcgtgaatgactgcga.....gaagaagaagaagaaaaa 1298

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 442306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPRO.spool/US09502945/runat_14032003_101101_19301/app_query.fasta_1.10979
-DB=PublishedApplications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09502945 -ECGN_1_1_60 -runat_14032003_101101_19301
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pcp:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	401	18.7	434	10	US-09-124-280A-3
2	313	14.6	611	10	US-09-216-393-81
3	279	13.0	369	9	US-09-820-843A-95
4	277.5	13.0	213	12	US-10-001-843-178
					Sequence 3, Appl1
					Sequence 81, Appl
					Sequence 95, Appl
					Sequence 178, App

5	260.5	12.2	197	9	US-10-101-487-51	Sequence 51, Appl
6	256.5	12.0	179	9	US-10-101-487-107	Sequence 107, App
7	256.5	12.0	197	9	US-10-101-487-114	Sequence 114, App
8	256.5	12.0	1005	10	US-09-925-301-1335	Sequence 1335, Ap
9	255.5	11.9	180	9	US-10-101-487-116	Sequence 116, Appl
10	251.5	11.8	176	9	US-10-101-487-56	Sequence 56, Appl
11	251	11.7	130	10	US-09-864-761-35536	Sequence 35536, A
12	250	11.7	186	9	US-10-101-487-44	Sequence 44, Appl
13	250	11.7	198	9	US-10-101-487-42	Sequence 42, Appl
14	250	11.7	200	9	US-10-101-487-53	Sequence 53, Appl
15	248	11.6	179	9	US-10-101-487-46	Sequence 46, Appl
16	248	11.6	240	9	US-10-101-487-75	Sequence 75, Appl
17	248	11.6	350	9	US-10-101-487-58	Sequence 58, Appl
18	247.5	11.6	376	10	US-09-925-301-1399	Sequence 1399, Ap
19	247	11.5	176	9	US-10-101-487-70	Sequence 70, Appl
20	247	11.5	187	9	US-10-101-487-50	Sequence 50, Appl
21	247	11.5	191	9	US-10-101-487-81	Sequence 81, Appl
22	246	11.5	177	9	US-10-101-487-48	Sequence 48, Appl
23	246	11.5	177	9	US-10-101-487-115	Sequence 115, App
24	246	11.5	181	9	US-10-101-487-45	Sequence 45, Appl
25	245.5	11.5	174	9	US-10-101-487-72	Sequence 72, Appl
26	245.5	11.5	175	9	US-10-101-487-57	Sequence 57, Appl
27	240.5	11.2	358	9	US-10-001-835-173	Sequence 173, App
28	239	11.2	617	10	US-09-864-761-36182	Sequence 36182, A
29	234.5	11.0	432	10	US-09-840-787-47	Sequence 47, Appl
30	234	10.8	242	10	US-09-864-761-36180	Sequence 36180, A
31	221.5	10.4	87	10	US-09-864-761-39625	Sequence 39625, A
32	220	10.3	665	9	US-09-820-843A-107	Sequence 107, App
33	218.5	10.2	89	10	US-09-864-761-35241	Sequence 35241, A
34	217.5	10.2	86	10	US-09-864-761-34671	Sequence 34671, A
35	213	10.0	1175	10	US-09-771-161A-224	Sequence 224, App
36	213	10.0	1175	10	US-09-771-161A-225	Sequence 225, App
37	213	10.0	1175	10	US-09-771-161A-226	Sequence 226, App
38	213	10.0	1233	9	US-09-291-417-13	Sequence 89, Appl
39	213	10.0	1239	9	US-09-291-417-13	Sequence 13, Appl
40	212.5	9.9	452	10	US-09-881-752A-268	Sequence 268, App
41	212	9.9	243	9	US-10-001-835-169	Sequence 169, App
42	210	9.8	259	9	US-09-925-299-821	Sequence 821, App
43	210	9.8	259	10	US-09-925-299-821	Sequence 821, App
44	209	9.8	257	10	US-09-922-261-188	Sequence 188, App
45	209	9.8	272	10	US-09-922-261-186	Sequence 186, App

ALIGNMENTS

RESULT 1
US-09-124-280A-3
Sequence 3, Application US/09124280A
Patent No. US20020034520A1
GENERAL INFORMATION:
APPLICANT: POTRO, Massimo
TITLE OF INVENTION: VACCINES FOR PREVENTION OF GRAM-
NEGATIVE BACTERIAL INFECTIONS AND
ENDOTOXIN RELATED DISEASE
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124,280A
FILING DATE: July 29,1998
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 434 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-09-124-280A-3

Alignment Scores:

Pred. No.: 5,81e-26 Length: 434
Score: 401.00 Matches: 79
Percent Similarity: 73.3% Conservative: 86
Best Local Similarity: 35.1% Mismatches: 56
Query Match: 18.74% Indels: 4
DB: 10 Gaps: 0

US-09-502-945-3 (1-1298) x US-09-124-280A-3 (1-434)

QY 630 GAAAGGCCTCTCTCGAGCCATGAGCGGCAAGGCAAGATGCGAGCAGAGCTT 689
Db 1 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 20
QY 690 CAGAGGCTCGGCTCTGAGCTCTACTATGANGCCCAAGCTCGAAGAGAGAAATC 749
Db 21 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
QY 750 NAAAGTTAAAGATCACAAAGTCGTGAAGAAAGGAAAGCCCAAGAACCCCTAAAGAG 809
Db 41 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 60
QY 810 TTTGACAGCTCGGAGGTAAATCCAGCTGCCGCTACTAGAAAGCAAGAAAGAGGAA 869
Db 61 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
QY 870 AGAAGGAGGAGGAG 929
Db 81 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100
QY 930 AGAAAGAGGAA 989
Db 100 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 120
QY 990 GAGGAGAGAAACAGAACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
Db 120 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 139
QY 1050 AAAAACTNAAAG 1109
Db 140 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 159
QY 1110 AAGAA 1169
Db 160 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 179
QY 1170 AAGAA 1229
Db 180 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 199
QY 1230 GAA 1289
Db 199 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 219
QY 1290 AAGGAAA 1296
Db 219 sLysLys 221
RESULT 2

US-09-216-393-81
Sequence 81, Application US/09216393
Patent No. US20010014447A1
GENERAL INFORMATION:
APPLICANT: Milhausen, Michael James
TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
FILE REFERENCE: TX-1-C2
CURRENT APPLICATION NUMBER: US/09/216,393
CURRENT FILING DATE: 1998-12-18
EARLIER APPLICATION NUMBER: 08/994,825
EARLIER FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 364
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 81
LENGTH: 611
TYPE: PRT
ORGANISM: Toxoplasma gondii
US-09-216-393-81

Alignment Scores:

Pred. No.: 1.75e-18 Length: 611
Score: 313.00 Matches: 100
Percent Similarity: 45.52% Conservative: 103
Best Local Similarity: 22.42% Mismatches: 154
Query Match: 14.63% Indels: 89
DB: 10 Gaps: 11

US-09-502-945-3 (1-1298) x US-09-216-393-81 (1-611)

QY 107 GATGAGGGGCAATGATGGAGAGAGAAAGCATCNAAGCTTCTCGAA----- 154
Db 130 AsnAspGlyAspArgGluGlyAspGluLysGluThrArgAspValGluAspGluGlyGlu 149
QY 155 -----GCAATCAGTTCCCTGTGATGGAAGAAATAGGCGGAAATGGCTCAN 199
Db 150 ThrArgArgSerSerPheAlaGluGlnThrGlyAsnGluArgThrGluMetArgThr 169
QY 200 AGG----- 202
Db 170 ArgHisGlyGlyAspGluGlyTrpThrSerLysSerAsnArgPheAlaPheAlaCysPro 189
QY 203 -----TCTGAGGCTAGTCTCGAAGTGTCTCAGGTCACAGTTC 232
Db 190 ArgPheSerLysSerAspValCysCysSerProGlnAlaArgLeuSerLeuProGluGln 209
QY 233 AATGTCAGTTCTGAAGGATCAGGAGAAAGCTGTCTGCAGACTCTGCTTGAGCCTGTT 292
Db 210 SerLeuGlySerSerProSerSerProIleSerValThrAsnAspValTyAlaLeuPhe 229
QY 293 AAAACTTCATCTCTTTTGGCCACTCTGAAAAAGCAACTGAGTAGAGTCANATCAAAGAAN 352
Db 230 AspSerSerAlaSer----- 234
QY 353 ACAGTGGAGTTACCTCTGAACAAAGAGAGATGAAACGGATCCACAGAGAAATAGCATCA 412
Db 235 -----ProLeuHisAlaGlyGluLeuSerSerLeuProGlyAlaValSerAla 250
QY 413 ATAAACGC-----ACAAGTCTCTCAAAATGGGACCTGTGCTCCTGAAGAACGGCA 466
Db 251 SerGluArgLeuThrAlaProAlaGluLeuGlyProSerAlaSerSerAlaCysLeu 270
QY 467 GCAGAGCAGCTGGTGTTCCTCGAAGAGAGGAGCGCAGCATTTGCTCCCATTTGAACA 526
Db 271 SerValSerCysGly-----ProGlyGluMet-SerProThrAlaAspThrThrArgH 288
QY 527 TGTGCTCAGTGGCTGGAAGGCAAGAACCTCCCTGGAGCAGGAGAAATTTTCAACCTCCTCA 586
Db 288 sAspAlaGluGluArgGluArgAlaGluGluLys-----G1 303
QY 587 TAAGAACAGAGCAGCAGTGCAGAGACCCCTTTACTGACCCCTGTGGAAGAGGCTCTCTCCG 646
Db 303 uArgGluArgGln-----GluGluGluGluArgGluArg 315

[illegible]

Db 556 spGluGluArgProArgArgLeuGlyAspGluAspArgGluProSerLeuArgProA 576
 QY 1280 AGAAGAAAGA 1290
 Db 576 spAspAspArg 579
 RESULT 9
 US-10-101-487-116
 ; Sequence 116, Application US/10101487
 ; Patent No. US20020169125A1
 ; GENERAL INFORMATION:
 ; APPLICANT: LEUNG, DAVID W.
 ; APPLICANT: BERGMAN, PHILIP A.
 ; APPLICANT: LOFOUIST, ALAN
 ; APPLICANT: PIETZ, GREGORY E.
 ; APPLICANT: TOMPKINS, CHRISTOPHER K.
 ; APPLICANT: WAGGONER JR., DAVID W.
 ; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
 ; FILE REFERENCE: 077319/0329
 ; CURRENT APPLICATION NUMBER: US/10/101,487
 ; PRIOR FILING DATE: 2002-03-20
 ; PRIOR APPLICATION NUMBER: 60/277,705
 ; PRIOR FILING DATE: 2001-03-21
 ; NUMBER OF SEQ ID NOS: 116
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 116
 ; LENGTH: 180
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
 US-10-101-487-116
 Alignment Scores:
 Pred. No.: 9,49e-14 Length: 180
 Score: 255.50 Matches: 65
 Percent Similarity: 61.66% Conservative: 54
 Best Local Similarity: 33.68% Mismatches: 48
 Query Match: 11.94% Indels: 26
 DB: 9 Gaps: 1
 US-09-502-945-3 (1-1298) x US-10-101-487-116 (1-180)
 QY 728 GGCCTCAAGAGAGAAACAAATCNAAGTTAAAGTATCACAAGTCTGTAAGAAAGAA 787
 Db 1 GlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 20
 QY 788 GGCCTCAAG 847
 Db 21 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 26
 QY 848 AGAAGAACCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 907
 Db 27 -----GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 39
 QY 908 GAAAGAAAG 967
 Db 40 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 59
 QY 968 AAAAGGAA 1027
 Db 59 luGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 78
 QY 1028 GAATAAGAA 1087
 Db 79 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 98
 QY 1088 AAGAGGAA 1147
 Db 99 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 118
 QY 1148 AGAAGAA 1207

Db 119 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 138
 QY 1208 AGGAGAAAGAA 1267
 Db 138 luGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 158
 QY 1268 ATAAGNAGAA 1296
 Db 158 luGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 167
 RESULT 10
 US-10-101-487-56
 ; Sequence 56, Application US/10101487
 ; Patent No. US20020169125A1
 ; GENERAL INFORMATION:
 ; APPLICANT: LEUNG, DAVID W.
 ; APPLICANT: BERGMAN, PHILIP A.
 ; APPLICANT: LOFOUIST, ALAN
 ; APPLICANT: PIETZ, GREGORY E.
 ; APPLICANT: TOMPKINS, CHRISTOPHER K.
 ; APPLICANT: WAGGONER JR., DAVID W.
 ; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
 ; FILE REFERENCE: 077319/0329
 ; CURRENT APPLICATION NUMBER: US/10/101,487
 ; PRIOR FILING DATE: 2002-03-20
 ; PRIOR APPLICATION NUMBER: 60/277,705
 ; PRIOR FILING DATE: 2001-03-21
 ; NUMBER OF SEQ ID NOS: 116
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 56
 ; LENGTH: 176
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
 US-10-101-487-56
 Alignment Scores:
 Pred. No.: 2,06e-13 Length: 176
 Score: 251.50 Matches: 65
 Percent Similarity: 62.69% Conservative: 56
 Best Local Similarity: 33.68% Mismatches: 54
 Query Match: 11.75% Indels: 18
 DB: 9 Gaps: 1
 US-09-502-945-3 (1-1298) x US-10-101-487-56 (1-176)
 QY 728 GGCCTCAAGAA 787
 Db 1 GlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 20
 QY 788 GGCCTCAAGAA 847
 Db 21 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 27
 QY 848 AGAAGAACCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 907
 Db 28 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 47
 QY 908 GAAAGAAAGAA 967
 Db 48 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 67
 QY 968 AAAGAA 1027
 Db 67 luGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 86
 QY 1028 GAATAAGAA 1087
 Db 87 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 106

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:32:38 ; Search time 23.1214 Seconds
(without alignments)
10793.685 Million cell updates/sec

Title: US-09-502-945-3

Perfect score: 2140

Sequence: 1 ggctgctgaatgactgcga.....gaagaagaagaagaaaaa 1298

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model
-Q=/cgn2_1/USPTO-spool/US09502945/runat_14032003_101059_19127/app_query.fasta_1.10979
-DB=PIR_73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_263 -runat_14032003_101059_19127 -NCPU=6 -ICPU=3
-NO_XLPYX -NO_MMAL -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_73:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	322.5	15.1	699	2 E84565	hypothetical prote
2	312.5	14.6	1701	2 T09127	probable erythrocy
3	307.5	14.4	1178	2 S78475	mannosylphosphoryl
4	298	13.9	1898	1 A45973	trichohyalin - hum
5	289.5	13.5	1407	1 S28589	trichohyalin - rab
6	287.5	13.4	335	2 T33457	hypothetical prote
7	285	13.3	1166	2 H86341	hypothetical prote
8	280	13.1	385	2 T19201	hypothetical prote
9	279	13.0	369	2 H71321	conserved hypothet
10	275	12.9	451	2 G70241	hypothetical prote
11	274.5	12.8	390	2 T34137	hypothetical prote
12	272	12.7	647	2 T29523	hypothetical prote
13	269	12.6	1240	2 S52734	hypothetical prote
14	267.5	12.5	1390	2 S51364	sperm tail-specifi

15	266	12.4	1052	1 A44937	kinetoplast-associ
16	264.5	12.4	312	2 T25994	hypothetical prote
17	264.5	12.4	810	2 T44430	protein PV100 (imp
18	262.5	12.3	877	2 T50591	class II INCENP pr
19	259	12.1	1432	2 B85431	trichohyalin like
20	259	12.1	1549	1 A40691	trichohyalin - she
21	258.5	12.1	771	1 A33430	h-caldesmon - chic
22	257.5	12.0	678	2 A54514	glutamic acid-rich
23	255.5	11.9	1085	2 S62516	hypothetical coile
24	254	11.8	899	2 S49634	hypothetical prote
25	253.5	11.8	729	2 T50989	hypothetical prote
26	251	11.7	929	2 T38948	hypothetical coile
27	250.5	11.7	522	2 C96608	hypothetical prote
28	249.5	11.7	849	2 S00030	neurofilament trip
29	249	11.6	1359	2 T34036	hypothetical prote
30	248	11.6	1344	2 T42637	hypothetical prote
31	247.5	11.6	484	2 A40988	54K arginine-rich
32	247.5	11.6	2962	2 T19756	hypothetical prote
33	246.5	11.5	301	2 T33068	hypothetical prote
34	246	11.5	793	1 JH0628	hypothetical prote
35	245	11.4	845	2 A45869	caldesmon - human
36	244	11.4	839	2 T50590	neurofilament trip
37	243.5	11.4	1403	2 T11583	class I INCENP pro
38	242.5	11.3	737	2 T15597	probable translati
39	242.5	11.3	1017	2 T15598	hypothetical prote
40	241.5	11.3	980	2 E71806	hypothetical prote
41	240	11.2	1877	2 T21861	hypothetical prote
42	239	11.2	1020	1 QFHUH	neurofilament trip
43	238.5	11.1	409	2 E86336	hypothetical prote
44	238.5	11.1	447	2 S52391	centrosomin B mo
45	235.5	11.0	762	2 G88436	protein T04A8.13

ALIGNMENTS

RESULT 1

E84565
hypothetical protein At2g18540 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C/Accession: E84565
R/Lin, X.: Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Cross-references: GB:AE002093; NID:g4218005; PIDN:AAD12213.1; GSPDB:GN00139
A/Accession: E84565
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-699 <STO>
A/Cross-references: GB:AE002093; NID:g4218005; PIDN:AAD12213.1; GSPDB:GN00139
C/Genetics:
A/Gene: At2g18540
A/Map position: 2

Alignment Scores:	Pred. No.:	Length:
Score:	1.97e-15	Matches: 699
Percent Similarity:	322.50	Conservative: 119
Best Local Similarity:	43.66%	Mismatches: 91
Query Match:	24.74%	Indels: 138
DB:	15.07%	Gaps: 133
	2	17

US-09-502-945-3 (1-1298) x E84565 (1-699)

QY	2	GCCTGCTGAATGACTGCGAACCCGGCTTCGAGAGAGCGCTTCCTGCTTTGAGCCANCAAGAA 61
Db	292	AlacysculleisSerilleValLeuGluGlyMetValArgValValAsnGlnGln 311
QY	62	GAAGTACGGATTTGCCAAAGACTACCTCTTTCAGTGAGAGT-----GAAGATGAGGGG 115
Db	312	SerLeuSerSerCysLysAsnAsp---ArgLysSerGluSerPheMetValGluGluGly 330

```
QY 116 GAC----- 118
|||
Db 331 AspValPheValProLysPheHisProMetAlaGlnMetSerPheGluAsnSerSer 350
QY 119 -----AATGATGGAGAGAGAAAGCATCNAAGCTTCTGGAAGCA 157
|||
Db 351 PheValPheMetGlyPheSerThrSerAlaLysThrAsnHisProGlnPheLeuValGly 370
QY 158 ATCAGTCTCCCTTGATGGAAAGATAGCGGAAATGGCTGGANAGTCTGAGGCTAGTGTG 217
|||
Db 371 GlnSerSerVal-----LeuLysValLeuAspArgAspValValAlaVal 385
QY 218 AAGGTGTCTAGAGTCAATGTCAGTCTCTGAAGGATCAGGAGAAAAGCTGCTTCACAGAT 277
|||
Db 386 -----SerPheAsnLeuSerAsnGlu-----ThrIleLysGly 396
QY 278 CTGCTTGAGCCCTGTAAACTTCA-----TCTTCTTTGGCCACTCTGAAAAGCA 328
|||
Db 397 LeuLeuLysAlaGlnLysGluSerValIlePheGluCysAlaSerCysAlaGluGlyGlu 416
QY 329 CTGAGTGTAGTTCANATCAAAAGAACAGTGGAGTTACCTCTGAACAAAGAGAGATTGAA 388
|||
Db 417 LeuSerLysLeuMetArgGluIleGluLysArgLysArgGluGluGluIleGlu 436
QY 389 CGGATCCACAGAGATPAGCATTTCAATPAAACGCACAAAGTCTCTCCAAATGGGACCTGT 448
|||
Db 437 ArgArgArgLysGluGluGluAlaArgLysArgGluGluAlaLysArg----- 453
QY 449 CGTCTGTAGAACCGGACGAGAGAGAGCTGTTTCCCTGGAGAAAGAGAGCCAGC 508
|||
Db 454 ArgGluGluGluAlaLysArg----- 462
QY 509 CATGTCTCCATGAACATGTGTCTAGTGTCTGGAAGCAAGAACTCCCTCGAGCAGGA 568
|||
Db 462 ----- 462
QY 569 AATTTTCAACCTCTCCATGAAGAACAGCAGCAGTGTGACAGACCTTTTACTGACCCCT 628
|||
Db 463 -----GluGluGluThrGluArgLysLysArgGlu----- 473
QY 629 GGAAGAGGCTCTCTCGAGCCATAGCTAGAGAGGCAAGATCGCAGCAGCAGCT 688
|||
Db 474 -----GluGluGluAlaArgLysArgGluGluGluArgLysArgGlu 487
QY 689 TCAGAGGGCTGGGCTCTGCACTGCTATGANGCCAAAGGCTCGAAGAGAGAAAGAAAT 748
|||
Db 488 GluGlu-----GluAlaLysArgArgGluGluGlu 497
QY 749 CNAAGTTAAAGTATCACAAGTCTGTGAAGAAAGGCAAGAAAGCCCTAAAGAA 808
|||
Db 498 ArgLysLysArgGluGluAlaGluGlnAlaArgLysArgGluGluArgGluLys 517
QY 809 GTTTGAGCAGCTCGGAAGTTAATCCAGCTGCCACTAGAAAGCAGAAAGAAAGAGA 868
|||
Db 518 Glu-GluGluMetAlaLys-----LysArgGluGluGlu 528
QY 869 AAGAAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 928
|||
Db 528 uArgGlnArgLysGluArgGluValGluArgLysArgGluGluArgGluGlnGluArgLys 548
QY 929 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 979
|||
Db 548 ArgArgGluGluGluAlaArgLysArgGluGluArgLysArgGluGluGluMetal 568
QY 980 AAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1033
|||
Db 568 aLysArgArgGluGluGlnArgGlnArgLysGluArgGluGluValGluArgLysIleArg 588
QY 1034 GAAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075
|||
Db 588 gGluGluGlnGluArgLysArgGluGluMetalAlaLysArgArgGluGlnGluArgGlu 608
```

```
QY 1076 GAAGAAGAAAGAGAGAGAACTNAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAA 1135
|||
Db 608 nLysLysGluArgGluMetGluArgLysLysArgGluGluGluAlaArgLysArgGlu 628
QY 1136 TAAGAACACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1186
|||
Db 628 uGluGluMetAlaLysIleArgGluGluGluArgGluArgLysGluArgGluAspValGlu 648
QY 1187 GAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
|||
Db 648 uArgLysArgArgGluGluGluAlaMetArgArgGluGluArgLysArgGluGluGlu 668
QY 1238 AAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
|||
Db 668 uAlaAlaLysArgAlaGluGluGluArgArgLysLysGluGluGluGluGluArg 687

RESULT 2
T09127
probable erythrocyte-binding protein MAEBL - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T09127
R:Kappe, S.H.I.; Noe, A.R.; Fraser, T.S.; Blair, P.L.; Adams, J.H.
Proc.Natl. Acad. Sci. U.S.A. 95, 1230-1235, 1998
A:Title: A family of chimeric erythrocyte binding proteins of malaria parasites.
A:Reference number: Z16577; MUID:98115903; PMID:9448314
A:Accession: T09127
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1701 <KAP>
A:Cross-references: EMBL:AF031886; NID:G2947227; PID:G2947228
A:Experimental source: subspecies yoelii; strain YM
C:Genetics:
A:Gene: maebl
A:Introns: 62/1; 1648/1; 1674/2; 1697/1
C:Keywords: alternative splicing; cell binding; erythrocyte invasion

Alignment Scores:
Pred. No.: 1,01e-14 Length: 1701
Score: 312.50 Matches: 99
Percent Similarity: 52.17% Conservative: 69
Best Local Similarity: 30.75% Mismatches: 114
Query Match: 14.60% Indels: 40
DB: 2 Gaps: 12

US-09-502-945-3 (1-1298) x T09127 (1-1701)
QY 417 AACGACAAAGTCTCTCCAA-----TGGGACCTGTCTC 452
|||
Db 1098 AsnAlaGluIleIleArgLysPheGluGluAlaGlnLysAlaAlaIleAlaLysLysAla 1117
QY 453 CTGAAGAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 512
|||
Db 1118 GluGluGluArgLysLysAlaGluAlaValLysLysAlaGluGluGluArgLysArgIle 1137
QY 513 -----GCTCCATTGAACATGTCTAGTGGCTGGAAGGCAAGCAACCTCCCTGGAG 563
|||
Db 1138 GluAlaGluLysLysAlaGluGluArgLysArgIleGluAlaGluLysLysAlaGlu 1157
QY 564 CAGGAATTTTCAACCTCTCCATGAAGAACAGCAGCAGCAGCAGCAGCAGCAGCAGC 623
|||
Db 1158 GluGluArgLysArgIleGluAlaGluLysLysAlaGluGluGluArgLysLysIleGlu 1177
QY 624 CCGTGGAAAGCGCTCTCTCGAGCCATGAGCCCTAGAGAGAGAGAGAGAGAGAGAGAG 683
|||
Db 1178 AlaAlaLysLysAlaGluGluArgLysArgIleGluGluAlaLysLysAlaGluGlu 1197
QY 684 GAGCTTCAGAGCGCTCGGCTCTGCAGCTCTACTATGANGCCAGCGCTCGAGAGAGAG 743
|||
Db 1198 GluArgLysLysIleGluAlaLys-----LysAlaGluGluGluArg 1212
QY 744 AAAATCNAAAGTTAAAGATATCATCAAAAGTGTGGAAGAGAGAGAGAGAGAGAGAG 797
|||
```

Db 1213 Lys-----LysAlaGluAlaValLysLysLysAlaGluGluAlaLysLysLysAla 1228
 QY 798 ---GCCCTAAAGAGTTTGCAGCAGCTGCGGAGGTTAATCCAGCTGCC-----CCACTA 848
 Db 1229 GluAlaLysLysAlaGluGluArgLysLysLysLysAlaGluAlaLysLysLysAlaLeu 1248
 QY 849 GAAGACCAAGAAAG 908
 Db 1249 GluArgLysLysLysSerGlu-----AlaAlaLysLysLysAlaLeuGluArgLysLysLys 1266
 QY 909 AAAG 968
 Db 1267 AlaGluAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1286
 QY 969 AAAG 1028
 Db 1286 uGluGluLysLysLysAlaGluAlaLysLysLysLysLysLysLysLysLysLysLys 1306
 QY 1029 AATAAG 1085
 Db 1306 sLeuGluArgLysProGluArgLysLysLysLysLysLysLysLysLysLysLysLys 1326
 QY 1086 AGAAG 1137
 Db 1326 gLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1346
 QY 1138 ---AGAACNAG 1193
 Db 1346 gGluArgLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1366
 QY 1194 AGAGAA-----GGAAG 1232
 Db 1366 sGluLysLysAlaGluAlaLysLysLysLysLysLysLysLysLysLysLysLysLys 1386
 QY 1233 GAAGAAAGTAGAAGCGGAG 1292
 Db 1386 uLysLysAlaGluGluArgLysLysLysLysLysLysLysLysLysLysLysLysLys 1406
 QY 1293 GAA 1296
 Db 1406 gLys 1407
 RESULT 3
 S78475
 N:Altenate names: protein MNN4 - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C:Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 19-Apr-2002
 R:Odani, T.; Shimizu, Y.; Yoshifumi, J.
 submitted to the EMBL Data Library, January 1996
 A:Description: Cloning and Analysis of the MNN4 Gene Required for Phosphorylation of N-1
 A:Reference number: S78475
 A:Accession: S78475
 A:Molecule type: DNA
 A:Residues: 1-1178 <ODA>
 A:Cross-references: EMBL:D83006; NID:g1752735; PID:d1012343; PID:g1752736; MIPS:YKL201c
 R:Maia e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.; Guerreiro, P.; H
 submitted to the Protein Sequence Database, March 1994
 A:Reference number: S38024
 A:Accession: S38037
 A:Molecule type: DNA
 A:Residues: 121-249, 'ATLOQT', 255, 'MALLRD', 262-390, 'LRISSEN', 398-515, 'LG' <MAI>
 A:Cross-references: EMBL:Z28200; NID:g486355; PID:g486356
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S78475
 A:Note: this was assumed to be protein YKL200c
 A:Accession: S38038
 A:Molecule type: DNA
 A:Residues: 640-1178 <MAW>
 A:Cross-references: EMBL:Z28201; NID:g486357; PID:g486359; MIPS:YKL201c
 A:Experimental source: strain S288C

A:Note: this sequence has been revised in reference S78475
 A:Note: this was assumed to be the complete sequence of protein YKL201c
 C:Genetics:
 A:Gene: SGD:MNN4; MNN4
 A:Cross-references: MIPS:YKL201c; SGD:S0001684
 A:Map position: 11L
 C:Function:
 A:Description: required for phosphorylation of N-linked oligosaccharides
 C:Keywords: transmembrane protein
 F:28-44/Domain: transmembrane #status predicted <TM>
 Alignment Scores: 2,3e-14 Length: 1178
 Score: 307.50 Matches: 69
 Percent Similarity: 69.94% Conservative: 52
 Best Local Similarity: 39.88% Mismatches: 37
 Query Match: 14.37% Indels: 15
 DB: 2 Gaps: 4
 US-09-502-945-3 (1-1298) x S78475 (1-1178)
 QY 792 AGAAGCCCTAAAGAGTTTGCAGCAGCTGCGGAGGTTAATCCAGCTGCCACTAGAA 851
 Db 1011 ArgLysGlyIleGlnMetPheAspLys-----AspProIleIleValTyrGlu 1026
 QY 852 GAA-----CGAAGAAAG 899
 Db 1027 AspTyrAlaTyrAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1046
 QY 900 GAGAAGAAAG 959
 Db 1047 GluGluGlu-LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1066
 QY 960 GGAGAAAGAAAG 1019
 Db 1066 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1085
 QY 1020 -----GAGAGGAG 1073
 Db 1086 GluGluGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1105
 QY 1074 AGAAG 1133
 Db 1106 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1124
 QY 1134 AATAAGACACNAG 1193
 Db 1125 AsnGluAspGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1144
 QY 1194 AGAAG 1253
 Db 1145 AsnLysLysAsnGluAspGluLysLysLysLysLysLysLysLysLysLysLysLys 1164
 QY 1254 AAAGAAAGAAAGTATAAG 1284
 Db 1165 GluGluGlu-GluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1174
 RESULT 4
 A45973
 trichohyalin - human
 C:Species: Homo sapiens (man)
 C:Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
 C:Accession: A45973
 R:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
 J. Biol. Chem. 268, 12164-12176, 1993
 A:Title: The structure of human trichohyalin. Potential multiple roles as a functional
 ed (cross-linking) protein.
 A:Reference number: A45973; MUID:93280194; PMID:7685034
 A:Accession: A45973
 A:Molecule type: DNA
 A:Residues: 1-1898 <LEE>
 A:Cross-references: GB:L09190; NID:g292835; PIDN:AAA65582.1; PID:g292836
 A:Note: authors translated the codon AGG for residue 1714 as Pro

C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sheath. Covalent modifications to this protein include conversion of arginine to citrulline and C;Genetics:

A;Gene: GDB:THH
A;Cross-references: GDB:136223; OMIM:190370
A;Map position: 1q21-1q21
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F;49-81/Domain: calmodulin repeat homology <EF2>

Alignment Scores:
Pred. No.: 1.09e-13 Length: 1898
Score: 298.00 Matches: 96
Percent Similarity: 48.10% Conservative: 106
Best Local Similarity: 22.86% Mismatches: 136
Query Match: 13.93% Indels: 82
DB: 12

US-09-502-945-3 (1-1298) x A45973 (1-1898)

```
QY 77 CCAAAAGACTACCTCTGAGTGAGAGTGAAGATGAGGGGAGCAATGATGAGAGAGAAAG 136
Db 124 ProArgAspArgGlnLeuGluGlu-----GluProGlyGlnArgArg 137
QY 137 CATCNAAAGCTTCTGGAACCAATCAGTTCCCTTGATGGAAGAAATAGCGGAAATTTGGCT 196
Db 138 ArgGlnLysArgGlnGlu-----GlnGluArgGluLeuAla 149
QY 197 GANAGGTCTGAGGTAGTCTGAAGGTGTGAGAGTTCAGTTCAGTTCGAAGGATCAGGA 256
Db 150 GluGlyGluGlnSerGluLysGlnGlnGluArgLeuGluGlnArgAspArgGlnArgArg 169
QY 257 GAAAGCTGGTCTGCGAGATCTGCTGAGCTGTAAACTTCATCTCTTGGCCACT 316
Db 170 AspGluGluLeuTrpArgGlnArgGlnGluTrpGlnGluArgGluGluArgAlaGlu 189
QY 317 GFGAAAGCACTCAGTGTAGTGCANATCAAGAAAGACAGTGGAGTTACCT----- 367
Db 190 ---GluGluGlnLeuGlnSerCysLysGlyHisGluThrGluGluPheProAspGluGlu 208
QY 368 ---CTGAACAAGAGAGATTGAACGGATCCACAGAGATAGCATTCATTAACAGGCACA 424
Db 209 GlnLeuArgArgGluLeuLeuGluLeuGluLeuArgArgLysGlyArgGluGluLysGln-Gl 228
QY 425 AGTCCTCTCCAAATGGACCCCTGCTGCTGAGAACCGGAGGAGGAGGAGGAGGAGGAGG 484
Db 228 n-----GlnArgArgGluArgGlnArgGlnArgGlnArgGlnArgGlnArgGlnArg 238
QY 485 TCCCTGGAGAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 544
Db 238 eGlnGluGluGluGluLysGlu-----ThrValLeuArgLysGluGluGluLysLe 247
QY 545 GGCAGAGAACTCCCTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 604
Db 247 GlyArgGlu-----ThrValLeuArgLysGluGluGluLysLe 260
QY 605 GACAGACCTTTACTACCCCTGTGGAAGGCTCTCTCCAGGAGGAGGAGGAGGAGGAGGAG 664
Db 260 uGlnGluGlu-----GluProGlnArgGlnArgGlnLeuGlnGlnGluGluGlnGlnLe 278
QY 665 GGCAGAGATGCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 724
Db 278 uArgLysLeuGluGluGlnGluLeuArgArgGluArgGlnGluGluGlnGlnGlnGln 298
QY 725 CAAGGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 784
Db 298 nArgLeuArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 308
QY 785 AAAGGCCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 844
Db 309 -----LysGlnGluGluGluArgArgGlu----- 316
QY 845 ACTAGAAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 904
```

```
Db 317 -GlnGlnGluGluArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 336
QY 905 GAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 964
Db 336 luGlnGlnLeuArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 356
QY 965 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021
Db 356 luGluGluArgArgGluGlnGlnLeuArgArgGluGlnGluGluGluGluArgGluGln 376
QY 1022 GAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081
Db 376 InLeuArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 396
QY 1082 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
Db 396 euArgArgGluGlnGlnLeuArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGln 416
QY 1142 ACNAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1177
Db 416 rgGluGlnGlnLeuArgArgGluGlnGlnLeuArgArgGluGlnGlnLeuArgArgGlu 436
QY 1178 GAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
Db 436 InGlnLeuArgArgGluGlnGlnLeuArgArgGluGlnGluGluGluGluArgHisGlu 456
QY 1238 AAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1293
Db 456 ySHisGluGlnGlnArgArgGluGlnArgGluGlnArgGluGlnGlnGluGluArgArg 474
```

RESULT 5

S28589
trichohyalin - rabbit

C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999

C;Accession: S28589

R;Fietz, M.J.; Rogers, G.E.

submitted to the EMBL Data Library, December 1992

A;Description: Examination of the gene encoding rabbit trichohyalin.

A;Reference number: S28589

A;Accession: S28589

A;Molecule type: DNA

A;Residues: 1-1407 <FIE>

A;Cross-references: EMBL:Z19092; NID:g1746; PIDN:CAA79519.1; PID:g1747

C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root

Covalent modifications to this protein include conversion of arginine to citrulline a

C;Genetics:

A;Introns: 46/3

C;Superfamily: trichohyalin; calmodulin repeat homology

C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F;49-81/Domain: calmodulin repeat homology <EF2>

Alignment Scores:
Pred. No.: 4.42e-13 Length: 1407
Score: 289.50 Matches: 130
Percent Similarity: 44.04% Conservative: 99
Best Local Similarity: 25.00% Mismatches: 160
Query Match: 13.53% Indels: 131
DB: 20

US-09-502-945-3 (1-1298) x S28589 (1-1407)

```
QY 23 CGGCTTGACAGAGAGCCCTCTCGCTTTGAGCCAGGAGAGAGAGAGAGAGAGAGAGAGAG 82
Db 728 ArgArgGluGlnGlnLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 745
QY 83 GACTACCTCTTGTAGTGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 133
Db 746 GluGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 765
QY 134 -----AAGCATCNAAAGCTTCTGGAAGCAATCAGTTCCTTCCTGAGAGAGAT 181
```


Db 766 LeuArgArgGluArgAspArgLysPheArgGluGluGluGlnLeuLeuGlnGluArgGlu 785
 Qy 182 AGCGGAAATTCGGCTGAGAGGCTCAGGCTAGCTGAGAGTG---TCAGAGTTCAATGTC 238
 Db 786 GluGluArgLeu---ArgArgGlnGluArgGluArgLysLeuArgGluGluGlnLeu 804
 Qy 239 AGTTCTGAAGGATCAGGAGAAAGCTGGTCTCTGAGAGTCTGCTGAGCCTGTTAAACT 298
 Db 805 LeuGlnGluArgGluGluArgLeu----- 813
 Qy 299 TCATCTCTTTGGCCACCTGTGAAAAGCACTAGTAGTACAGTACATCAATCAATAAAA 358
 Db 814 -----ArgArgGlnGluArgGluArgLysLeuArgGluGlu 825
 Qy 359 GAGTTACCTCTGACAAAGAGAGATTGAACGGATCCACAGAGATACATCAATCAATAAAA 418
 Db 826 GluGlnLeuLeuGlnGlnGluArgGluGluArgLeuArgGln-----Glu 841
 Qy 419 CGCACAACTCTCTCAAAATGGGACCCCTGCTGCTGAAGAACCGCAGGAGAG----- 473
 Db 842 ArgGluArgLysLeu-ArgGluGluGlnLeuLeuArgGlnGluGlnGlnLeuArg 861
 Qy 474 ---CAGCTGCTTTTCCCTGGAGAAAGAGAGAGAGCCATTCCTCCATTTGACATGT 529
 Db 861 glnGluArgAlaArgLysLeuArgGluGluGlnLeuLeuArgGlnGluGlnGln 881
 Qy 530 GCTCAGTGGCTGGAGGCAAGAACTCCCTCGAGGAGGAGAAATTTCAACCTCCCTCCATAA 589
 Db 881 uLeuArgGlnGluArgAspArgLysLeuArgGluGlu-----GlnLeuLeuArgGln 899
 Qy 590 GAACAAGCAGCCAGTG-----ACAGACCCCTTTACT 619
 Db 899 nGluGlnGlnLeuLeuArgGlnGluArgAspArgLysLeuArgGluGlnGlnLeuLe 919
 Qy 620 GACCCCTGTGAAAAGGCTCTCTC-----CGAGCCATGAGCCTAGA 661
 Db 919 uGlnGluSerGluGlnGluArgGlnGluArgGlnGluArgGlnGluArgLysLeuArgGln 939
 Qy 662 AGAGCAAAAGATGCGACGA-----GCAGAGCTTCAGAGGCTCGGGCTCTGCAGTCTTA 715
 Db 939 uGlnLeuLeuArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 955
 Qy 716 CTATGANGCCAAAGCTCGAGAGAGAGAAATNAAAGTTAAAGTATACAAAGTCGT 775
 Db 956 -----ArgLysLeuArgGluGluGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGln 973
 Qy 776 GAAGAAAGGAAAGCCAAAGAGCCCTTAAAGAGTTTTCAGAGCTG----- 821
 Db 973 uArgGlnGlnGluArgAlaArgLysLeuArgGluGlnGlnGlnGlnGlnGlnGlnGln 993
 Qy 822 -----CGGAAGTTAAATCCACCTGCCCTAGAGACG 856
 Db 993 uGlnGluLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1013
 Qy 857 AAGAAAGAGGAAAGAGGAGGAGG----- 881
 Db 1013 uArgGluGlnGluArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1033
 Qy 882 -----AGAAGAGAGAACAGAGAGAGAA 907
 Db 1033 glnLeuArgArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1053
 Qy 908 GAAAGAGAGGAG 963
 Db 1053 gLeuGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1073
 Qy 964 -----AAGAAAG 1012
 Db 1073 spArgLysPheArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1093
 Qy 1013 GAAGAGAGAGAGAG----- 1026
 Db 1093 rgArgGluArgAspArgLysPheArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1113

Qy 1027 --AGAATAAG 1084
 Db 1113 luArgLeuArgArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1133
 Qy 1085 AAG 1138
 Db 1133 rgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1153
 Qy 1139 GGAACNAG 1192
 Db 1153 euGlnGluSerGlu---GluGluArgLeuArgArgGlnGlnGlnGlnGlnGlnGln 1172
 Qy 1193 AAG 1252
 Db 1172 luGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1192
 Qy 1253 GAAAGAGAGAGAGAG-----TATAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
 Db 1192 rgLysLeuArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1210
 RESULT 6
 T33457
 hypothetical protein F36H12.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33457
 R:Blanchard, M.; Bradshaw, H.; Stellyes, L.
 submitted to the EMBL Data Library, July 1998
 A:Description: The sequence of C. elegans cosmid F36H12.
 A:Reference number: 221346
 A:Accession: T33457
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-335 <BLA>
 A:Cross-references: EMBL:AF078790; PIDN:AAC26930.1; GSPDB:GMO0022; CESP:F36H12.3
 A:Experimental source: strain Bristol N2; clone F36H12
 C:Genetics:
 A:Gene: CRSP:F36H12.3
 A:Map position: 4
 A:Introns: 32/3; 227/1; 270/2

Alignment Scores:

Pred. No.: 6,18e-13 Length: 335
 Score: 287.50 Matches: 70
 Percent Similarity: 64.21% Conservative: 52
 Best Local Similarity: 36.84% Mismatches: 48
 Query Match: 13.43% Indels: 20
 DB: 2 Gaps: 3

US-09-502-945-3 (1-1298) x T33457 (1-335)

Qy 740 GAAGAAATCNAAGTTTAAAGTATCACAAGTCGTGAAGAAAGAGAGAGAGAGAGAG 799
 Db 51 GluGluLysGluLysSerLysLysSerLysLysSerLysLysSerLysLysSerLysLys 70
 Qy 800 CCTAAAGAGTTTGAGCAGCTGCGGAAGTTAATCCAGTCCGCCACTAGAGAGAGAG 859
 Db 71 LysLysGlu-----LysSerLysLys 77
 Qy 860 AAAAG 919
 Db 78 SerGluGlyLysLysSerLysLysSerLysLysSerLysLysSerLysLysSerLysLys 96
 Qy 920 GAGAAGGAG 979
 Db 97 ---Lys-GluAspLysLysGluLysLysGluLysLysGluLysLysGluLysLysLys 115
 Qy 980 AAAAG 1039
 Db 115 spGluLysLysAspGluLysLysAspGluLysLysAspGluLysLysSerGluAspLys 134
 Qy 1040 AAG 1093

Db	21	lleylspheasnproasnleuInThrProLysalaGluProGluProMetArgPro	40
QY	473	GCAGCTGGTTTTTCCCTGGAGAAAGAGAGCCAGCCATTGCTCCCATTTGAACATGTGCT	532
Db	41	AspThrThrThrSerSerGlySerArgProAlaSerSerLysSerAsn	56
QY	533	CAGTGGCTGGGAAGCAGAACTCCCTGGAGCAGGAATAATTTTCAACCTCTCCA	586
Db	57	TyrHisAsnGluProProProProAlaAsnProProProLeuLys	72
QY	587	-----TAAGAACAGCAGCC	601
Db	73	PheArgPheLysAsnLeuPheAsnLeuGlyAspGlnAspValLysLysGluGluProSer	92
QY	602	AGTGACAGACCTTTACTGACCCCTGTGGAAAAGGCTCTCTCCGAGCCATGAGCCTAGA	661
Db	93	SerMetThrProGluSerSerArgProGlySerSerLeuGlu	106
QY	662	AGAGCAAGATCGCAGCAGCAGCTTCAGAGGGCTCGGGCTCCAGTCTCTACTATGA	721
Db	107	-----ThrProSerSerSerSerSerLysHis-HisHisHisHisHis	121
QY	722	NGCCAAGGCTCGAGACAGAGAAATCNAAAGTTAAAGTATCAACAAGTCTGNAAGA	781
Db	121	sLysLysGluArgLysAspLysGlu	135
QY	782	AGGAAAGCCCAAGAAAGCCCTAAAGAGTTTGACAGCTCGGAAGTTTAATCCAGCTGC	841
Db	135	sAspArgGluHisArg	140
QY	842	CGCACTAGACAGCAAGAAAGAGAAAGAGGAGGAGAGAAAGAAAGAAACAAGAA	901
Db	141	-AspArgGluLysGluArgGluArgAspGluArgLysGluArgGluArgGluGlnLysGln	160
QY	902	GAACAAGAAAGAGGAGAGGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG	961
Db	160	uLysGluArgGluAspAlaAlaArg	175
QY	962	AGAAAGAAAGGAGAGAAAGAAAGAAAGAGAGAGAAAGAAAGTAAAGAAAGAA	1021
Db	175	aGluMetAspAlaLysArgValalaGluGluGluGluGluArgLysGluLysGluLys	195
QY	1022	GAGCAAGAAATAGAGAAAGAGAAAGAAAGAAAGTAAAGAGAGAGAGAGAGAA	1081
Db	195	sArgArgGluGluLysLysLysGlnLysGluLeuLeuLysGluLysGluArgSerGluArg	215
QY	1082	GAAAAGAGAGGAGAACTNAGAAGAAAGAAAGAGGAGAGAGAAAGAGAGAGAAATAGGA	1141
Db	215	gLysGluLysGluArgGluLeuGluArgGluLysGluLysSerArgGluLysGluArgGln	235
QY	1142	ACNAGAAGAGAGAGAAAGAAATAGAGAGAGAGAGAGAAAGAAAGAGAGAGAGAG	1201
Db	235	uLysGluArgGluLysGluArgGluLysGluArgGluLysGluArgGluLys	253
QY	1202	GAAGAAGGAGAAAGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1261
Db	253	nLysGluArgGluLysGluArgGluLysGluArgGluLysGluLysLysArgGluGlu	272
QY	1262	GAAGTATAG	1295
Db	273	-----GluAlaArgArgLysLysGluGlu	280

RESULT 9
H71321
conserved hypothetical protein TP0470 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: H71321
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dod-
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterb-
rger, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:title: Complete genome sequence of Treponema pallidum, the syphilis spiro-

A:Reference number: A71250; MUID:983327745; PMID:9665876
A:Accession: H71321
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-369 <COL>
A:Cross-references: GB:AE001223; GB:AE00520; NID:g3322745; PID:g332
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0470

Alignment Scores:	
Pred. No.:	2,49e-12
Score:	279.00
Percent Similarity:	48.43%
Best Local Similarity:	30.19%
Query Match:	13.04%
DB:	2
	Gaps: 14
	Indels: 68
	Mismatches: 97
	Conservative: 58
	Matches: 96
	Length: 369

Alignment Scores:		
Pred. No.:	2,49E-12	Length:
Score:	279.00	Matches:
Percent Similarity:	48.43%	Conservative:
Best Local Similarity:	30.19%	Mismatches:
Query Match:	13.04%	Indels:
DB:	2	Gaps:
		14
US-09-502-945-3 (1-1298)	x H71321 (1-369)	

US-09-502-945-3 (1-1298) x H71321 (1-369)

QY	450	GTCTGTAAGAACCGGAGGAGCAGCTGGTGTTCCTCCCTCGAGAAA-----GAG	500
Db	48	LeuLeuSerLeuGlnAlaGlnAlaIleGlyProLeuHisLysAlaAlaGlnGln	67
QY	501	GAGCCAGCC-----ATTGCTCCCATGTACATGTGCTCAGTGGC-----	539
Db	68	LysProAlaHisProLysAlaAlaLeuTyrLeuGlyMetAlaTyrLeuGlnThrGlyArg	87
QY	540	-----TGGAGGGCAAGAACCTCCCTCGAGCAGAA-----	569
Db	88	TyrThrGlnAlaIleGlnTrpLeuGlnAsnProValHisSerGlnGluTyrAlaHis	107
QY	570	-----ATTTCACCTCTCCATTAAGAACAAGCAGCAGCATGTG	605
Db	108	LeuTyrAlaTyrAsnLeuGlyAsnValTyrPheValGlnHisArgTyrGluGlu	125
QY	606	ACAGACCCTTTACTGACCCCTGTGGAAAAGGCTCTCTCCGAGCCATGAGCTAGAACAG	665
Db	126	-----AlaGlnHisAlaTyrGluGlnAlaLeuAlaLeuLysHis	138
QY	666	-----CCAAAGATGCACGACGAGCTTCAGAGGCGCTCGGCGCTCGCAGTCC	713
Db	139	AspTyrProProAlaLeuLeuAsnArgAlaAsnThrAlaMetLysArg-----GlnAla	156
QY	714	TACTATGANGCCAGGCTCGAAGAGAGAAGAAATCNAAGTTAAAGATATCACAAATC	773
Db	157	TyrAlaHisAlaLeuAla-----AspTyrLysLysTyr	167
QY	774	GTCAAGAAAGGAAAGGCCATAAAGAGTTTTCAGACAGCTCGGAAGGTTAAT	833
Db	168	ValSerGlnAsnProThr-----AlaSerGlnHisTyrGluValGlnArgMetIle	184
QY	834	CCAGCTGCCGACTAGAGAA-----CGAAGAAAGAGGAAAGAGGAGGAGGAGAAAG	887
Db	185	-----AlaAlaLeuGlnTrpLeuGlnArgLysGluAlaGluGluAlaArgArgLys	202
QY	888	ARGAAGAACACGAGAGAGAAGAGNAGAGGAGGAGGAGAGAGAGAGAGAGAGAGAG	947
Db	203	GluAlaGluGluAlaArgLysGluAlaGluGluAlaArgArgLysGluAlaGluGlu	222
QY	948	---GAAAGAGAGAACGACGAGAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG	1004
Db	223	AlaArgArgLysGluAlaGlu-----GluAlaArgArgLysGluAlaGluGluAla	239
QY	1005	AGAACTTAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1064
Db	240	ArgArgLysGluAlaGluGluAlaArgLysGluAlaGluGluAlaArgArgLysGlu	259
QY	1065	GAAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1123
Db	260	AlaGlu-GluAlaArgArgLysGluAlaGluGluAlaArgArgLysGluAlaGluAla	279
QY	1124	AGAAAGAGAAGTAAGGAACNAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1183

Db 279 aaRgARGLySGluLaGluGluLaalaaRGaRGLySGluLaGluGluLaalaaRGaRGLySGl 299
OY 1184 AAAGAAGAAAGGAAGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGTAAG 1243
: : : : : ||||| | : : : : :
Db 299 ualaGluGluLaalaaRGaRGLySGluLaGluGluLaalaaRGaRGLySGluLaGluGluLa 319
: : : : : ||||| | : : : : :
OY 1244 AAAGCGGAAGAAAGAAAGTAATRAAGAGGAAGGAAGGAAGGAAGGAAGGA 1295
: : : : : ||||| | : : : : :
Db 319 aaRgARGLySGluLaGlu-----GluLaalaaRGaRGLySGluLaGlu 333

RESULT 10
G70241
hypothetical protein BB116 - Lyme disease spirochete plasmid I/ip28-4
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C;Accession: G70241
R;Praser, C.M.; Castens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White-
son, D.; Peterson, J.; Keriavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
J.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: G70241
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-451 <KLE>
A;Cross-references: GB:AE000789; MID:g2690079; PIDN:AC66203.1; PID:g2690100; TIGR:BB116
A;Experimental source: strain B31
C;Genetics:
A;Genome: plasmid

```

Alignment Scores:
Pred. No.:      4.8e-12      Length:      451
Score:          275.00      Matches:      76
Percent Similarity: 60.96%      Conservative: 77
Best Local Similarity: 30.28%      Mismatches: 70
Query Match:     12.85%      Indels:      28
DB:              2          Gaps:       7

US-09-502-945-3 (1-1298) x G70241 (1-451)

QY 573 TTCAACCTCCCTCCATAAGAAC--AAGCAGCAGTGACAGACCCCTTTTACTGACCCCTGTG 629
      |||||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 21 PheASnIleAspGlnLysAspIleLysTyrPro-----ProThr 33
      |||||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 630 GAAAGGCGCTCTCCGAGCCATGAGCTAGCAAGAGGCAAAAGATCGCAGCAGAGCTT 689
      |||||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 34 GluLysSerArgProLysThr-----GluSerLysGlnLysGluSerLysPro 50
      |||||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 690 CAGAGGGCTCGGGCTCTGCATCTCTACTATGANGCCAGAGGTCGAGAGAGAGAAATC 749
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 51 LysThrGluGluGluLeuLys-----LysLysGlnGlnGluGluGluLeu 65
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 750 NAAAGTTAAAGTATCACAAAGTCTGTAAGAAAGCAAGGCCAAGAGCCCTAAAA--- 806
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 66 LysLysLysGlnGlnGluGluGluLeuLysLysLysGlnGlnGluGluGluLysLys 85
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 807 -----GAGTTTGAGCAGCTCGGAAGGTTAATCCAGCTGCGCAGTACAAGAGCAAGA 860
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 86 LysGlnGlnGluGluGluLeuLysLysLysGlnGlnGluGluGluLeuLysLysGln 105
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 861 AAAGAGGAAAGAGGAGGAGGAGAAAGAAAGCAACAAAGGAGAGAGAAAGAAAGAGGG 920
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 106 GlnGluGluGluLeuLysLysLysGlnGlnGluGluGluLeuLysLysGlnGlnGlu 125
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 921 AGAAGGAGAAAGAAAGAGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGGA 980
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 126 GluGluLeuLysLysLysGlnGlnGluGluGluLeuLysLysLysGlnGlnGluGlu 145
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 981 AAAGGAAAGAGAGAGAGAGAGAGAAAGAACTAAGAAAGAGAGAGAGAAATAAGAGGA 1040
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

[illegible]

Mol. Biochem. Parasitol. 40, 233-244, 1990

A: Title: Cloning and characterization of a gene coding for a protein (KAP) associated with kinetoplasts in *Leishmania* spp.

A: Reference number: A44937; PMID: 1694571

A: Accession: A44937

A: Molecule type: DNA

A: Residues: 1-1052 <ON>

A: Cross-references: EMBL:M25364; NID:g162141; PID:AA30209.1; PID:g162142

C: Comment: This protein was detected only in kinetoplasts of replicative stages of the parasite.

C: Comment: This protein is translocated to the mitochondrion without the loss of an amino acid.

C: Superfamily: kinetoplast-associated protein

C: Keywords: duplication; mitochondrion; tandem repeat

F: 424-563, 582-705, 724-862/Region: 9-residue repeats (A-A-R-K-Q-A-E-E-E)

Alignment Scores:		
Pred. No.:	2,09e-11	1052
Score:	266.00	109
Percent Similarity:	40.48%	Matches: 78
Best Local Similarity:	23.59%	Conservative: 186
Query Match:	12.43%	Mismatches: 89
DB:	1	Indels: 89
		Gaps: 13
US-09-502-945-3 (1-1298) x A44937 (1-1052)		

119	QY	A	NTGATGGAGAGAAAGCATCNAAGCTTCTGGAAGCAATCAGTTCCTT-----GAT	172
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
284	Db	Ser	SerGluGluGluArgHisGluGlnIleAlaGluLysPheSerGluAlaProLysGlu	303
173	QY	G	AAAGAATAGCGCGAAATTCGCTGANAAGTCTCAG-----	208
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
304	Db	Asp	VallLysArgAsnProPheAlaMetMetSerLysLeuMetLysHisSerArgHisPhe	323
209	QY	G	CTAGTCTGAAGGTCTCAGAGTTCAAATGCTCAGTCTCTGAAGGATCAGGA-----	256
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
324	Db	Gln	ThrCysLysThrAlaAlaGluSerProSerAsnThrGlyGluAsnAlaGlnThrGln	343
257	QY	-----	GAAGAAGTGGTCTTTCGAGATCTGCTTGAGCCTGTATAAAGTTCATCT-----	304
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
344	Db	Thr	LysGluLysValGluValProGluValIleValThrGlyAlaThrSerGluIleAla	363
305	QY	-----	TCTTTGGCCACTGTCAAAAACCAACTGAGT	334
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
364	Db	Gln	GlyThrProValCysAsnAspGlnProAsnLeuProGlnSerSerAspGlyIleArg	383
335	QY	A	GAGTCANATCAAGAAANACAGTGGAGTTACTCTGAACAAAGAGAGATTGAACGGATC	394
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
384	Db	Gly	ValGluAsnSerLysThrAlaGluMetPro-----ArgHisThrIleLysLys---	400
395	QY	C	ACAGAGAATAGCATTCATAATAAAACGACAAAGTCCCTCTCCAAATGGGACCTGCTGCTCT	454
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
401	Db	-----	HisGluLysLysLysThrLysLysSerLeuMetLysArgLysLeuMet	416
455	QY	G	AAGAACCGGACGAGCAGACAGCTGGTTTTCCTCTGGAGAAAGAGGAGCCAGCATTCG	514
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
417	Db	Lys	GluMetAlaGluLysProAlaAlaLeuLysGlnAlaGlu-GluGluAlaAlaLeuLys	436
515	QY	T	CCCATCAACATGTGCTCAGTGGCTGGAAG-----	545
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
436	Db	s	glnAlaGluGluAlaAlaArgLysGlnAlaGluGluAlaAlaArgLysGlnAl	456
546	QY	-----	GCAAGAACTCCCTGGAGCA	565
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
456	Db	a	gLuGluAlaAlaArgLysGlnAlaGluGluAlaAlaArgLysGlnAlaGluG	476
566	QY	G	GAATTTTCACTCTCCATAGAACACAGCAGCCATGTGACACCCCTTTACTGACCCC	625
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
476	Db	u	GluAla-----AlaArgLysGlnAlaGluGluAlaAlaAlaArgLysG	491
626	QY	T	GTGAAAGGCCCTCTCTCCGACCATGAGCCTACAAGAGGCAAAAGATGCGCAGGACGA	685
		:::	:::: :::: :::: :::: :::: :::: :::: :::: ::::	
491	Db	n	AlaGluGluAlaAlaArgLysGlnAlaGluGluAlaAlaArgLysGlnAlaG	511
686	QY	G	CTTCAGAGGGCTCGGGCTCTGCAGTCTCTACTATGANGCAAGGCTCGAAGAGAGA	745

Search completed: March 21, 2003, 13:05:13
Job time : 42.1214 secs

GenCore version 5.1.4.p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 11:21:37 ; Search time 12.2332 Seconds
(without alignments)
8801.668 Million cell updates/sec

Title: US-09-502-945-3
Perfect score: 2140
Sequence: 1 ggcgtgctgaatgactgcga.....gaagaagaagaagaagaaaa 1298

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=fragnet_n2p_model -DEV=xlip
-Q/cgn2_1/USPTO.spool/US09502945/runat_14032003_101058_19100/app_query.fasta_1.10979
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945.qcgn_1.1113/runat_14032003_101058_19100 -NCPU=6 -ICPU=3
-NO_XLPY -NO_MMAP -LARGEDUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	307.5	14.4	1178	1	MN4_YEAST	P36044	saccharomyc
2	298	13.9	1898	1	TRHY_HUMAN	Q07283	homo sapien
3	289.5	13.5	1407	1	TRHY_RABIT	P37709	oryctolagus
4	269	12.6	1240	1	YNJ1_YEAST	P53935	saccharomyc
5	264.5	12.4	1391	1	MST2_DROHY	Q08696	drosophilla
6	262.5	12.3	877	1	INCE_CHICK	P53352	gallus gall
7	259	12.1	1549	1	TRHY_SHEEP	P22793	ovis aries
8	248.5	12.1	771	1	CALD_CHICK	P12957	gallus gall
9	7.5	12.0	678	1	GARP_PLAFF	P13816	plasmodium
10	5	12.0	1382	1	IF3A_HUMAN	Q14553	homo sapien
11	5	11.9	1085	1	YAF3A_SCHPO	Q09863	schizosacch
12	4	11.9	899	1	YMJ3_YEAST	Q04500	saccharomyc
13	11.7	9.9	929	1	YDM6_SCHPO	P87137	schizosacch
14	11.6	8.8	848	1	NFM_MOUSE	P08553	mus musculus
15	11.6	1359	1	1	ATRX_CAEEL	Q9u7e0	caenorhabdi
16	11.6	1344	1	1	IF3A_MOUSE	P23116	mus musculus
17	11.6	484	1	1	SFRB_HUMAN	Q05519	homo sapien
18	11.5	793	1	1	CALD_HUMAN	Q05682	homo sapien

19	244.5	11.4	845	1 NFM_RAT	P12839 rattus norv
20	243.5	11.4	1403	1 YDF3 SCHPO	Q10475 schizosacch
21	239	11.2	1020	1 NFM_HUMAN	P12036 homo sapien
22	235	11.0	705	1 TRDN_RABIT	Q28870 oryctolagus
23	234.5	11.0	488	1 CYL2_BOVIN	Q28092 bos taurus
24	234	10.9	1394	1 CNG4_BOVIN	Q28181 bos taurus
25	232	10.8	915	1 NFM_HUMAN	P07197 homo sapien
26	227.5	10.6	1002	1 IF2P_YEAST	P39730 saccharomyc
27	227	10.6	831	1 NFM_RAT	P16884 rattus norv
28	226.5	10.6	2492	1 ATRX_HUMAN	P46100 homo sapien
29	226	10.6	978	1 RA50_AQUAE	O67124 aquifex aeo
30	226	10.6	1220	1 IF2P_HUMAN	O60841 homo sapien
31	224.5	10.5	644	1 NFM_RABIT	P54938 oryctolagus
32	224	10.5	494	1 SFR4_HUMAN	Q08170 homo sapien
33	222.5	10.4	810	1 NFM_BOVIN	O77788 bos taurus
34	221.5	10.4	728	1 TRDN_HUMAN	Q13061 homo sapien
35	220.5	10.3	344	1 MST1_DROHY	O08695 drosophila
36	219.5	10.3	805	1 IF2_AQUAE	O67825 aquifex aeo
37	218	10.2	471	1 RUL7_XENLA	P09406 xenopus lae
38	214.5	10.0	700	1 TRDN_CANFA	P82179 canis famli
39	214.5	10.0	911	1 CAFA_MOUSE	Q9qwf0 mus musculu
40	213	10.0	829	1 T2_XENLA	P41512 xenopus lae
41	212	9.9	185	1 T2_MOUSE	O06666 mus musculu
42	212	9.9	348	1 CYL2_HUMAN	Q14093 homo sapien
43	212	9.9	448	1 RUL7_DROME	P17133 drosophila
44	211.5	9.9	767	1 TOPI_CRIGR	Q07050 cricetus
45	211.5	9.9	2468	1 MAPE_HUMAN	P46821 homo sapien

ALIGNMENTS

RESULT 1

MN4_YEAST
ID MN4_YEAST STANDARD; PRT; 1178 AA.
AC P36044; P36043; P89095;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MN4 protein.
GN MN4 OR YKL200C/YKL201C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=97175967; PubMed=9023541;
RA Odani T., Shimma Y.-I., Tanaka A., Jigami Y.;
RT "Cloning and analysis of the MN4 gene required for phosphorylation
of N-linked oligosaccharides in Saccharomyces cerevisiae.";
RL Glycobiology 6:805-810(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Maia e Silva A., Bossier P., Villela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: MAY FUNCTION AS A POSITIVE REGULATOR FOR
MANNOSYLPHOSPHATE TRANSFERASE. IS REQUIRED TO MEDIATE
MANNOSYLPHOSPHATE TRANSFER IN BOTH THE CORE AND OUTER CHAIN
PORTIONS OF N-LINKED OLIGOSACCHARIDES.
CC -! SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -! SIMILARITY: TO YEAST YJR061W.
CC -! CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

[illegible]


```

Db 557 AspAspIleLeuAspGlyAsp----- 563
QY 590 GAACAAGCAGCAGTACAGACACCTTACTGACCCCTGTGGAAGAGGCTCTCTCCGAGC 649
Db 564 -----GluAspGluProGluGluGluAspGluAsnGluGlyAspAspGluGlu 579
QY 650 CATGAGCCTAGAGAGCAAGATGCGAGCAGCAGAGCTTCAGAGGCTCGGCGCTGCA 709
Db 580 AspThrTyAspSerGlyLeuAspGluThrAspArgLeuGluGluGlyArgLysLeuL 599
QY 710 GTCCTACTATGCGCAGGCTCGAAGAGAGAGAAATCNAAGTTAAAGTATACACAA 769
Db 599 eGlnIleAlaIleThrLysLeuLeuGlnSerArgIleMetAlaSer-----TyrHisG 617
QY 770 AGTCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 829
Db 617 uLysGlnAlaAspAsnArgLeuLysLeuGlnGluGluGlu----- 632
QY 830 TAATCCAGCTGCCGCACTAGAGAAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAA 889
Db 633 -----GluGluLysArgLysLysLysLysLysLysLysLysLysLysLysLys 646
QY 890 GAAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 949
Db 646 nLysLysArgGluLysGluLysGluLysLysLysLysLysLysLysLysLysLys 666
QY 950 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1009
Db 666 uLys-ArgLys-----ArgGluGluGluGluGluGluGluGluGluGluGlu 685
QY 1010 TAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1069
Db 685 LuMetArgArgGluAlaGlnArgLysLysValGluGluAlaLysArgLysLysAsp 705
QY 1070 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
Db 705 LuGluArgLysArgGluGluGluGluGlnArgGluGluGluGluGluGluGln 725
QY 1118 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1177
Db 725 rGlyGlnLysGluGluLysLysArgLysArgGluGluGluLysLysArgLysArg 745
QY 1178 GAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
Db 745 lNysArgLeuGluGlnGluLysLeuGlnLysGluLysGluGluGluGluGluGln 765
QY 1238 AAGTA---GAAGCGGAGAGAGAGAGAGAGAGAGATATAGAGAGAGAGAA 1281
Db 765 euIleAlaGluAspAlaLeuArgLysGlnLysLeuAsnGluGluGln 780

```

RESULT 5

MST2_DROHY

```

ID MST2_DROHY STANDARD; PRT; 1391 AA.
AC Q08696;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein mst101(2).
GN MST101(2).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=95045538; PubMed=7957199;
RA Neesen J., Padmanabhan S., Buenemann H.;
RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid
RT motif representing the major component of the sperm-tail-specific
RT axoneme-associated protein family Dhmst101 form extended
RT alpha-helical rods within the extremely elongated spermatzoa of

```

```

RT Drosophila hydei.";
RL Eur. J. Biochem. 225:1089-1095(1994).
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
CC SPERMATIDS.
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
CC REPEATS.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

```

CC -----
DR EMBL; X73481; CAA51876.1; .
DR PIR; S34154; S34154.
DR FlyBase; FBgn020733; Dhyd\mst101(2).
KW Sperm; Repeat; Multigene family; Polymorphism.
FT DOMAIN 332 1368
FT [KR]-K-X-C-X-X-K-X-K-X-K-X-X-E.
SQ SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48878 CRC64;

```

Alignment Scores:

```

Pred. No.: 1,4e-09 Length: 1391
Score: 264.50 Matches: 112
Percent Similarity: 42.86% Conservative: 71
Best Local Similarity: 26.23% Mismatches: 175
Query Match: 12.36% Indels: 69
DB: 1 Gaps: 11

```

US-09-502-945-3 (1-1298) x MST2_DROHY (1-1391)

```

QY 128 GAGAGAAAGCATCNAAGCTTCTGGAAGCA-----ATCAGTTCCTCTGATGAAAG 178
Db 265 LysArgValIleAlaGluMetCysGluAlaAlaGlyTyrvLysSerSerGluProLys 284
QY 179 AATAGCGGGAATTTGGCTGAGAGGTCTGAGGCTAGTCTGAAGGTCGATGTCATGTC 238
Db 285 LysLysGlyLysLysLysLysAsnAspGluLysLysGluLysGluGluGluGlu 304
QY 239 AGTTCTGAGGATCAGGAGAAAGCTGGTCTGCAGATCTCTTGAGGCTCTTAAA--- 295
Db 305 LeuLysGluGlnAlaGluGluGluAlaLysIleArgGlyValValLysLysLys 324
QY 295 ----- 295
Db 325 LysCysLysGluLysAlaLeuLysLysLysCysLysAspLeuLysMetLysGlu 344
QY 296 -----ACTTCATCTCTTTGGCCACTGTGAAAGAAAGCACTAGTAGCTCANATCA 346
Db 345 GluAlaGluLysLysLysCysAlaAlaLeuAlaLysGlnLysGluGluAspGluLys 364
QY 347 AAGAANACAGTGGAGTTACCTCTGAACAAGAAAGAGATT---GAACGGATCCACAGAGA 403
Db 365 LysAlaCysLysGluLeuAlaLysLysLysLysGluAlaAspGluLysLysCysGlu 384
QY 404 TAGCATTTCAATAAAGCGCAAGTCTCTCCAAATGGGACCTGTCGTCCTCGAAGAACCG 463
Db 385 GluAlaAlaAsnLysGluLysLysAlaAlaGluLysLysLysCysGluLysAlaAla 404
QY 464 GCAGGAGAGCAGCTGGTTTTTCCCTCGAGAAAGAGAGAGCCAGCCATTCCTCCATTGA 523
Db 404 sGluArgLysGluAlaAlaGluLysLysLysCysGluGluAlaAla----- 419
QY 524 ACATGTGCTCAGTGGCTGGAAGGCAAGAACTCCCTCGGAGCAGGAAATTTTCAACCTCT 583
Db 420 -----LysLysGluGluAlaAlaGluArgLysLysCysGluGlu 434

```


use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL; 218361; CAA79165.1; -
CC EMBL; X51695; CAA35992.1; -
CC PIR; A34209; A34209.
CC PIR; S32633; S32633.
CC PIR; A40691; A40691.
CC HSP; P02633; IIG5.
CC InterPro; IPR001751; CABP_S100.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00036; efhand; 1.
CC Pfam; PF01023; S_100; 1.
CC ProDom; PD003407; CABP_S100; 1.
CC PROSITE; PS00018; EF_HAND; 1.
CC PROSITE; PS00303; S100_CBP; FALSE_NEG.
CC Keratinization; Repeat; Calcium-binding.
KW DOMAIN 1 91
FT CA_BIND 22 33
FT CA_BIND 62 73
FT DOMAIN 413 832
FT REPEAT 413 448
FT REPEAT 449 476
FT REPEAT 477 504
FT REPEAT 505 532
FT REPEAT 533 560
FT REPEAT 561 588
FT REPEAT 589 616
FT REPEAT 617 644
FT REPEAT 645 678
FT REPEAT 679 706
FT REPEAT 707 742
FT REPEAT 743 771
FT REPEAT 772 796
FT REPEAT 797 832
FT REPEAT 833 861
FT DOMAIN 938 1507
FT REPEAT 938 961
FT REPEAT 962 985
FT REPEAT 986 1021
FT REPEAT 1022 1044
FT REPEAT 1045 1067
FT REPEAT 1068 1090
FT REPEAT 1091 1121
FT REPEAT 1122 1144
FT REPEAT 1145 1167
FT REPEAT 1168 1197
FT REPEAT 1198 1227
FT REPEAT 1228 1250
FT REPEAT 1251 1273
FT REPEAT 1274 1296
FT REPEAT 1297 1319
FT REPEAT 1320 1342
FT REPEAT 1343 1368
FT REPEAT 1369 1391
FT REPEAT 1392 1416
FT REPEAT 1417 1439
FT REPEAT 1440 1461
FT REPEAT 1462 1484
FT REPEAT 1485 1507
FT VARIANT 1145 1197
FT VARIANT 1251 1273
FT CONFLICT 1399 1399
SQ SEQUENCE 1549 AA; 201173 MW; E72FB9FF1326E54E CRC64;

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; 218361; CAA79165.1; -
EMBL; X51695; CAA35992.1; -
PIR; A34209; A34209.
PIR; S32633; S32633.
PIR; A40691; A40691.
HSP; P02633; IIG5.
InterPro; IPR001751; CABP_S100.
InterPro; IPR002048; EF-hand.
Pfam; PF00036; efhand; 1.
Pfam; PF01023; S_100; 1.
ProDom; PD003407; CABP_S100; 1.
PROSITE; PS00018; EF_HAND; 1.
PROSITE; PS00303; S100_CBP; FALSE_NEG.
Keratinization; Repeat; Calcium-binding.
KW DOMAIN 1 91
FT CA_BIND 22 33
FT CA_BIND 62 73
FT DOMAIN 413 832
FT REPEAT 413 448
FT REPEAT 449 476
FT REPEAT 477 504
FT REPEAT 505 532
FT REPEAT 533 560
FT REPEAT 561 588
FT REPEAT 589 616
FT REPEAT 617 644
FT REPEAT 645 678
FT REPEAT 679 706
FT REPEAT 707 742
FT REPEAT 743 771
FT REPEAT 772 796
FT REPEAT 797 832
FT REPEAT 833 861
FT DOMAIN 938 1507
FT REPEAT 938 961
FT REPEAT 962 985
FT REPEAT 986 1021
FT REPEAT 1022 1044
FT REPEAT 1045 1067
FT REPEAT 1068 1090
FT REPEAT 1091 1121
FT REPEAT 1122 1144
FT REPEAT 1145 1167
FT REPEAT 1168 1197
FT REPEAT 1198 1227
FT REPEAT 1228 1250
FT REPEAT 1251 1273
FT REPEAT 1274 1296
FT REPEAT 1297 1319
FT REPEAT 1320 1342
FT REPEAT 1343 1368
FT REPEAT 1369 1391
FT REPEAT 1392 1416
FT REPEAT 1417 1439
FT REPEAT 1440 1461
FT REPEAT 1462 1484
FT REPEAT 1485 1507
FT VARIANT 1145 1197
FT VARIANT 1251 1273
FT CONFLICT 1399 1399
SQ SEQUENCE 1549 AA; 201173 MW; E72FB9FF1326E54E CRC64;

Alignment Scores:
Pred No.: 3.05e-09
Score: 259.00
Percent Similarity: 43.81%
Best Local Similarity: 100
Query Match: 24.76%
Indels: 134
Gaps: 1
Length: 1549
Matches: 130
Conservative: 100
Mismatch: 161
Indels: 134
Gaps: 24

US-09-502-945-3 (1-1298) x TRHY_SHEEP (1-1549)
QY 47 TTGAGCCANCAAGAAAGACAGCGGATTTGCCAAAGACACCTCTTGAGTGAGAGTAA 106
Db 716 LeuGlnArgGlnGluGluArgLeuGlnArgGlnGluGlnGlnGlnGlnGlnGlu 735
QY 107 GATGAGGGGACAAATGATGAGAGAGAGAAACATCNAAG----- 145
Db 736 GluLysArgArgGlnValArgGluArgLysTyrLeuGluGluGlnGlnGluGlu 755
QY 146 ---CTTCTGCAAGCAATCAGTCCCTTCATGATGGAAGAAATAGCGGAAATTTGGCTGANAGG 202
Db 756 AspArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 775
QY 203 TCTGAGGCTAGTCTGAAGGTGTCAGAGTTCATGTCAGTTCAGTTCGAA-----GGATCA 253
Db 776 GluLysValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 795
QY 254 GGAGAAAGCTGTCCTTCGAGATCTGTCGAGCTGTTAAACATTCATCTTCTTTGGCC 313
Db 796 ArgGluArgGlnTyrArgGluGlu-----GluLeuLeuArgGluGluArgLeuHis 813
QY 314 ACTGTGAAAAGCAACTGACTAGA-----GTCANATCAAGAAGAACAGTGGAG 361
Db 814 ArgLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 833
QY 362 TTACCTCTGAACAAAGAGATTGAACGATC----- 394
Db 834 ArgGlnLeuGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 853
QY 395 -----CACAGAAATAGCATTCATTAACGACAGTCTCTCCAAA----- 437
Db 854 AspGlnHis-GlnAsnGluValArg---AsnSerArgValTyrSerLysHisArgGluAs 872
QY 438 -----TGGGACCTCTGCTCTGCTGAAGAACCGGACGACGACGAGCT 478
Db 872 nLysGluLysSerArgGlnLeuAspSerTyrPValArgGluSerGlnPheGlnGlnAs 892
QY 479 GGTTCCTTCCTCGAGAGAGAGGAGCCATTCCTCCATTCGATGACATGTGCTAGTGG 538
Db 892 pLeuArgProLeuGlnAspGluGlnGluGluLysArgGluArgGlu-----GlnG 909
QY 539 CTGAAGGCAAGAACT-----CCCTGGAGCAGGAAATTTTCAACCT 580
Db 909 uTyrArgSerArgGlnLysArgAspSerGlnPheProAlaGluGln-----Le 925
QY 581 CCTCCATAAGCAAGCAGCAGTGACA-----GA 610
Db 925 uLeuGluArgGluGlnGlnGlnGlnGlnGlnGlnGluArgArgAspArgLysPheArgGluGlu 945
QY 611 CCCTTTACTGACCCCTGTGGAAAAGGCTCTCTCCGAGCCATGAGCTAGAGAGGCAAA 670
Db 945 uGlnLeuLysGlyGlnArgGluGluLysIleArg-----TyrLeuGluGluAspArg 963
QY 671 GATGCCAGCAGAGAGCTTCAGAGGGCTCGGGCTCTGCTACTATGATGANGCCAAAGGC 730
Db 963 gLysPheArgGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 983
QY 731 TCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 790
Db 983 nGluArgAspArg-----LysPheArgGluGluLeuSerArgGlnGluArg 998
QY 791 CAAGAAAGCCCTAAAAGAGTTTGAGCAGCTGCGGAGAGTTAATCCAGCTGCCACATAGA 850
Db 998 gAspArgLysPheArgGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1010
QY 851 AGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902
Db 1010 nGluArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1030
QY 903 -----AGAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 943
Db 903 -----AGAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG

```



```
|||||
245 sGluGluAla-----GluLysGluArgGluLysLeuGluAlaGluLys 260
QY TCAGAGGCTGGGCTCCTCAGTCTACTATGANGC-----AAGGCTCGAAG 736
Db sGluArgLeuLysAlaGluGluLysLysAlaAlaGluGluLysGlnLysAlaGlu 280
QY AGAGAAGAAATNAAGTTAAAGTATCACAAGTCGTGAAGAAAGAAAGCCCAAGAA 796
Db uGluLysLysAlaAla-----GluGluArgGluArgAlaLysAlaGluGluLysAr 298
QY AGCCCTAAAGAGTTTACAGCAGCTCGGAGGTATATCAGCTCCGCGACTAGAGACG 856
Db gAlaAlaGluGluArgGluArg-----AlaLysAlaGluGluArg 312
QY AAGAAAGAGGAAAGAGAGGAGGAGAGAAAGAAAGAAAGAAAGAAAGAA 916
Db gLysAlaAlaGluGluArgGluArgAlaLysAlaGluGluArgLysAlaAlaGlu 332
QY AGGAGAGAGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 973
Db ArgAlaLysAlaGluGluArgGluArgLysAlaAlaGluGluArgAlaLysAl 349
QY AAGAGAAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATAA 1033
Db aGluGluGluArgLysAlaAlaGluGluArgAlaLysAlaGluGluArgLysAla 369
QY GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1090
Db aGluGluArgGluArgAlaLysAlaGluGluLysArgAlaAlaGluGluLysAla 389
QY AGGAAGAACTNAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1150
Db gLeuGluAlaGluLysLysLysLysLysLysLysMetGluGluLysAlaGlnGlu 409
QY AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
Db uLysAlaGlnAlaAsnLeuLeuArgLysGlnGluGluAsp-LysGluAlaLysValGlu 429
QY GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1246
Db lalLysLysGluSerLeuProGluLysLysLysLysLysLysLysLysLysLys 449
QY GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1281
Db spAsnLysAspLysGluLysAlaProLysGluGlu 460

RESULT 9
GARP_PLAFF
ID GARP_PLAFF STANDARD; PRT; 678 AA.
AC P13816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutamic acid-rich protein precursor.
GN GARP.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89040048; PubMed=2903445;
RA Triglia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.;
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
RT acid-rich protein (GARP).";
RL Mol. Biochem. Parasitol. 31:199-202(1988).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03998; AAA29605.1; -.
DR PIR; A54514; A54514.
KW Repeat; Malaria; Antigen; Signal.
FT SIGNAL 1
FT CHAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.
FT DOMAIN 120 164 15 X 3 AA TANDEM REPEATS OF K-K-X.
FT DOMAIN 372 416 9 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 417 441 5 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 576 604 POLY-GLU.
FT DOMAIN 605 653 7 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 654 663 POLY-GLU.
SQ SEQUENCE 678 AA; 80551 MW; 2A8F85608496EA9E CRC64;

Alignment Scores:
Pred. No.: 4.06e-09 Length: 678
Score: 257.50 Matches: 105
Percent Similarity: 39.91% Conservativeness: 77
Best Local Similarity: 23.03% Mismatches: 157
Query Match: 12.03% Indels: 117
DB: 1 Gaps: 14

US-09-502-945-3 (1-1298) x GARP_PLAFF (1-678)
QY 89 CTCCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 148
Db 49 LeuLeuAsnGluThrGluLeuGluLysAsnLysAspAsnSerLysSerGluThrLeu 68
QY 149 CTG-----CAAGCATCAGTCCCTGTGATGGAAGTGAAGTGAAGTGAAGTGAAG 187
Db 69 LeuLysGluLysAspGluLysAspValProThrThrSerAsnAspAsnLeuLys 88
QY 188 AATTTGGCTGANAGCTGAGCTAGTCTGAAGGTGTGAGAGTTCATATGCTAGTCTGAA 247
Db 89 AsnAlaHisAsnAsnGluLeuSerSerSerThrAspProThrAsnLeuLeuVal 108
QY 248 GGATCAGGAGAAAGCTGCTGCTGAGATGCTGAGCTGCTGAGCTGCTGAGCTGCTGCT 307
Db 109 AsnAspLysAspAsnGluAsnSerValAspLysLysLysLysLysLysLysLysHis 128
QY 308 TTGGCCACTGTCAAAAGCAACTGAGTGAAGTCAATCAAGCAAGCAAGTGAAGTACCT 367
Db 129 LysLysAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 148
QY 368 CTGAACAAGAAGAGATTGAACGGATCCACAGAGATAGCATTCAATAAAGCCAGAGT 427
Db 149 GluLysLysHisLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 168
QY 428 CCTCTC-----CAATGGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 475
Db 169 GluValMetSerLeuTyrLysThrGlyGlnHisLysProLysAsnAlaThr----- 185
QY 476 GCTGTTTTTCCCTGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
Db 186 -----GluHisGlyGluGluAsnLeuAspGlu-----Glu 195
QY 536 TGGCTGAAGGCAAGAACTCCCTGGAGCAGGAGAAATTTTCAACCTCTCCATCAAGAACAA 595
Db 196 MetValSerGluLeuAsnAsnAlaGlnGlyGlyLeu-----Leu 209
QY 596 GCAGCCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 652
Db 210 LeuSerSerProTyrGlnTyrArgGluGlnGlyGlyCysGlyLeuLeuSerValHis 229
QY 653 GAGCCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 712
Db 230 Glu----- 230
QY 713 CTACTATGANGCAAGGCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 772
Db 231 -----ThrSerAsnAspThrLysAspAsnAspLysGluAsnLysSerGluAsp 246
```



```
QY 299 TCATCTTCTTGGCCACTGTGAAAGCAACAGTGTAGTGTAGTCANATCAAGAAACAGTG 358
Db 634 Lys-----ThrValArgGluArgLeuGluGlnle-----643
QY 359 GAGTTACCTCTGAACAAAGAGATGTACGATCCACGAGAAATAGCATTAATAAA 418
Db 644 -----LysLysThrGlu-LeuGlyAlaLysAlaPheLysAspIle-----As 657
QY 419 CGCACAGTCTCTCAATGGACCTGTGTCCTGGAAGAACCGGACGACGACGAGCT 478
Db 657 pIleGluAspLeuGluLeuAspPheIleMetAlaLysGlnValGluGln-- 676
QY 479 GGTTCCTCCGAGAAAGAGGACGACCATGCTCCCATTTGAACATGTGCTCAGTGG 538
Db 677 -----LeuGluLysGluLysGluLeu-----684
QY 539 CTGGAAGGCAAGACTCCCTGAGCAGGAGAAATTTCAACCTCCTCCATAAGAACAGCA 598
Db 685 -----GlnGluArgLeuLysAsnGlnGluLysLysIleAspTyrPheGluArgAlaLysAr 703
QY 599 GCCAGTCACAGACCTTTACTG---ACCCCTGTGGAAAGGCCCTCTCTCGAGCCATGAG 655
Db 703 gLeuGluGluIleProLeuIleLysSerAlaTyrGluGluGlnArgIleLysAspMetAs 723
QY 656 CCTA-----GAAGAGCAAGATGCGACGACGACAGCTTCAGAGGCTCGGCG 703
Db 723 pLeuTrpGluGlnGlnGluGluArgIleThrThrMetGlnLeuGluArgGluLysAl 743
QY 704 TCTGCAGTCCTACTATGANGCCAGGCTCGAAGAGAGAGAAATCAATCAAGTTAAAGTA 763
Db 743 aLeuGluHisLysAsnArgMetSerArgMetLeuGluAspArgAspLeuPheValMetAr 763
QY 764 TCACAAATCGTGAAGAAAGAGGCAAGAGCCCTCAAGAGTTTTCAGAGCTGCG 823
Db 763 gLeuLysAlaArgGlnSerValTyrGluGluLysLeuLysGlnPheGluArgLe 783
QY 824 GAAGTTAATCCAGCTCCCTCCACTAGAACGACGAAAGAAAGAAAGAGAGGAGGAG 883
Db 783 uAlaGluGluArgHisAsnArgLeuGluGluArgLysArgGlnArgLysGluArgAr 803
QY 884 AAAGAAGCAAGCAAGGAGAGAGAGAGAGAGGAGGAGAGAGAGAGAGAGAGAGAG 943
Db 803 gIleThrTyrTyrArgGluLysGluGluGluGluGluGluGluGluGluGluGluGlu 823
QY 944 AGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1003
Db 823 uLysGluArgGluGluArgGluArg-----AlaG1 833
QY 1004 AAGAACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1063
Db 833 uArgAlaLysArgGluGluGluLeuArgGluTyrGlnGluArgValLysLysLeuGluG1 853
QY 1064 AGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1123
Db 853 uValGluArgLysLysArgGlnArg-GluLeuGluIleGluGluArgGluArgArg 873
QY 1124 AAGAAGAAAGA-----1134
Db 873 luGluGluArgArgLeuGlyAspSerSerLeuSerArgLysAspSerArgTrpGlyAspA 893
QY 1135 -----ATAGGAACNAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1176
Db 893 rgAspSerGluGlyThrTrpArgLysGlyProGluAlaAspSerGluTrpArgArgGlyP 913
QY 1177 -----AGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1219
Db 913 roProGluLysGluTrpArgArgGlyGluGlyArgGspGluAspArgSerHisArgArg 933
QY 1220 AAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1279
Db 933 spGluGluArgProArgArgLeuLysAspGspGluAspArgGluProSerLeuArgProA 953
```

```
QY 1280 AAGAAGAAGA 1290
Db 953 spAspAspArg 956
```

RESULT 11

```
YAF_A_SCHPO
ID YAF_A_SCHPO STANDARD; PRT; 1085 AA.
AC 009863;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C29E6.10c in chromosome I.
GN SPAC29E6.10c OR SPAC30.14c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA Holroyd S., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA James K., Mounie S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Mooney P., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Oliver K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Rutherford K., Skellern M., Squares R., Squares S., Stevens K.,
RA Skelton J., Simmonds M., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Taylor K., Taylor R.G., Tivey A., Tivey R., Robben J., Grymonprez B.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welte J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: TO YEAST YNL091W.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
Alignment Scores:
Pred. No.: 5,19e-09 Length: 1085
Score: 255.50 Matches: 104
Percent Similarity: 42.40% Conservative: 94
```

```
Best Local Similarity: 22.27%  Mismatches: 158
Query Match: 11.94%  Indels: 111
DB: 1  Gaps: 13
```

US-09-502-945-3 (1-1298) x YAFA_SCHPO (1-1085)

QY	59	GAAGAACTACGGATTTCGCAAAAGACTACTCCTCTTGAGTGAGAGTGAAGATGAGGGGAC	111
Db	325	GlUlnTyrAlaAsnIleGlnAAsnLeuAlaAsnThrGluSerValAsnAlaSerAsp	344
QY	119	AATGATGGAGAGAAAGCATCNAAGCTTCTGGAACATCAGTTCCTCTGATGGAAS	178
Db	345	GlUcLySerAsp	349
QY	179	AATAGCGGAAATTTGGCTGANAGTCTGAGGCTAGTCTGAAGGTGTGAGAGTTCATATGTC	238
Db	350	SerGlnLysGlyIleIleSerAspSerProLysLeuLeuSerIleProLeuAsnAsnVal	369
QY	239	AGTTCCTGAAGGATCAGGAGAAAGCTGGTCTCTGCAGATCTGCTGAGCTGTTAAACT	298
Db	370	ProSerLysSerLeuAsnAspIleThrGlnAspGluLeu	383
QY	299	TCATCTTCTTGGCCACTGTGAAAAGCACTCAGTAGAGTCANATCAAGAANACAGTG	358
Db	384	AsnSerSerAsnAlaaspValaspGluGluValIleGluThrThrSerLeuGluGluLys	403
QY	359	GAGTTACTCTGAACAAAGAGAGATTGACGGATCCACAGAAATAGCATTCATATAAA	418
Db	404	AsnValAsp--AsnGlnGluPheValThrSerIleSerAsnGlyAsnGlnThrLeuGlu	422
QY	419	CGACACA----AGTCTCTCCAANTGGGACCTGTGCTCTGCAAGAACCGCAGGCAG	472
Db	423	AspThrSerHisSerProGlnThrGlnProPheGlnProProTyrProSerLys---	441
QY	473	GCAGCTGGTTTTTCCCTGGAGAAAGAGGACCGACCATTTGCTCCCATTTGAACATGTGCT	532
Db	441	----	441
QY	533	CAGTGGCTGAAGGCAAGAACTCCCTCGGAGCAGGAAATTTCAACCTC-	581
Db	442	----AlaAspGluLysAsnSer-TyrHisSerAspLeuTyrAsnPheGlySerSerLe	459
QY	582	----CTCATTAAGAACAAGCAGCCAGTCAGACACCTTTTACCTGACCCCTGTGGAAAGGC	637
Db	459	uThrValLysGlyClyIleLeuThrValAlaLaspLeuLeuLysAsnAspGlyLysL	479
QY	638	CTCTCTCCGAGCATG---AGCCTAGAAGGCAAGATCGCAGCAGCAGAG-	686
Db	479	sPheIleGluMetMetGluGlnLeuAlaGluArgArgMetGlnArgGluAspAsnSerAs	499
QY	686	----	686
Db	499	nPheHisGluProGluLeuTyrGluSerGlyLeuGluTyrAspGluAspGluGluAs	519
QY	687	----CTTCAGAG	694
Db	519	pGluGluAspValaspGluAspGluLeuAspLeuMetThrAspGluGlnArgMetGluG	539
QY	695	GGCTCGGCTCTCAGTCTCTACTATGANGCCAAAGGTCTCGAAGAGAGAGAAATCNAAG	754
Db	539	uGlyArgArgMetPheGlnIlePheAlaAlaArgLeuPheGluGlnArgValLeuGlnAl	559
QY	755	TTAAAGATATCACAAAGTCTGTAAGAAAGGAGGCAAGCCCTAAAGAGTTTCA	814
Db	559	a-----TyrArgGluLysValalagInArgGlnAlaLysLeuLeuGluGluIleG	577
QY	815	GCAGCTCGCGAAGTTAATCCAGCTCCCGCATCTAGAAAGCAAGAAAGAGGAAGAG	874
Db	577	uGluGluAsnLysArgLysGlnGluArgGluLeuLysLysIleArgGluLysGluLysL	597
QY	875	GAGGAGGAGAAAGAGNAGAACAGGANGAAGAAAGAGGAGGAGGAGAGAGAA	934
Db	597	sArgAspLysLysGlnLeuLysLeuAlaLysGluGluGlu-----ArgGlnArg	615

Qy	935	AGAAGGAGAAGCAAAAGGAAGGAAGAAAGAAAAGGAAGGAGGAAAAAGAAAAGG	999
Dy	615	gglualagluargleuAlagluglnlaAlaGlnLysAlaLeu--GlualaLysArgGl	634
Qy	995	AGRAAGAAGAARACTYAGAAAGACAGAGAGAGGAATATAGAGGAAGAAGAAAAGAA	105
Dy	634	nGlugluaAargLysLysargGluGlnArgLeuLysArGluGlnGluLysLysGl	654
Qy	1053	-----AACTTAAACAAGAAAGAAAGAAAGGAAGGAAGGAAGGAAGAGGAG	1099
Dy	654	nGlnGluLeuGluArgGlnLysargGluGluGlnLysGlnLysGlnLysGluArgGluLys	674
Qy	1097	AACNTNAGAAGAAAGAGAGGAGAAAGAAAGAAAGAAATAAGAGAACNAGAAAGAGG	1155
Dy	674	sLeuLysLysGlnGlnGlnlualaaAsparGluLysMetAlaaGluGlnArgLeuAr	694
Qy	1157	AAGAAAGAATTAAGAGAGGAAAGAAAAGAAAGAAAGAAAGAAAGAAAGGAGAAA	1210
Dy	694	gGlugluGluGluLysArgfLeuGluGluArgLysArGrgGluLys-----	710
Qy	1217	AGGNAGAAAAAGCAGAAAGTAGAAGCGGRAGAAAGAAAGAAAGATATACAGAG	1276
Dy	711	----LeuAspLysGluGluGluArgArg-ArgGluLeuGluLysGluSerG	729
Qy	1277	AAGAAGAAGAAAGAAGG 1293	
Dy	729	LuGluLysGluArgArg 734	
RESULT 12			
ID	YMJ3 YEAST	STANDARD;	PRT; 899 AA.
AC	Q04500;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	15-JUL-1998 (Rel. 36, Last annotation update)		
DE	Hypothetical 103.0 kba protein in RAD10-PRS4 intergenic region.		
DN	YML093W.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
RX	NCBI_TaxID=4932;		
ON	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=S288c / AB972;		
RC	Gentiles S., Bowman S., Barrell B.G., Rajandream M.A.;		
RL	Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.		
CC	-I- SIMILARITY: TO C.ELSGANS C33G8.2 AND S.POMBE SPAC5747.06.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	tween the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for commen-		
CC	entities requires a license agreement (See http://www.isb-sib.ch/annou-		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; Z46660; CAAB6645.1; -		
DR	SGD; S0004558; YML093W.		
KW	Hypothetical protein; ATP-binding.		
FT	NP_BIND 260 267 ATP (POTENTIAL).		
FT	DOMAIN 162 168 POLY-SER.		
FT	DOMAIN 179 182 POLY-GLU.		
FT	DOMAIN 704 712 POLY-LYS.		
SQ	SEQUENCE 899 AA; 103023 MW; 0D4FC9D0ICB3CFE1 CRC64;		

Alignment Scores:	
Pred. No.:	6,54e-09
Score:	254.00
Percent Similarity:	41.03%
Best Local Similarity:	26.71%
Query Match:	11.87%
DB:	1
Length:	899
Matches:	135
Conservative:	67
Mismatches:	175
Indels:	101
Gaps:	22

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 295396; CAB08764.1; --
KW Hypothetical protein.
SQ SEQUENCE 929 AA; 105060 MW; 24D5A70AC9353E1A CRC64;

Alignment Scores:
Pred. No.: 1e-08 Length: 929
Score: 251.00 Matches: 118
Percent Similarity: 39.02% Conservative: 74
Best Local Similarity: 23.98% Mismatches: 183
Query Match: 11.73% Indels: 117
DB: 1 Gaps: 18

US-09-502-945-3 (1-1298) x YDM6_SCHPO (1-929)

QY	98	GAGAGTGAAGATGAGGGGACATGATGAGAGAGAGAGCATCNAAAGCTT-----	148
Db	209	GLuSerSerAspSerAspHisAspAspGlyGluAsnSerAspSerLysLeuAspAsnLeu	228
QY	149	-----CTGGAAGCA 157	
Db	229	ArgAsnTyrIleValSerLeuAsnGlnLysArgLysLysAspGluAlaAspAlaGluSer	248
QY	158	ATCAGTTCCCTTGATGAGAAAGATAGG-----CGAATTCGGCTGAN 199	
Db	249	ValLeuSerSerAspAsnAspSerIleGluIleSerIleLysValLysTyr 268	
QY	200	AGGTCTGAGGCTACTCTCAAGGTCTCAGATTCATGCTGAGGATCAGGAGAA 259	
Db	269	AspProHisGluThrAsnLysGluSerGluTyrAsnLeuLeu-----GlySerSerGlu 286	
QY	260	AAGCTGTC---CTGCAGATCTGCTGAGCCCTGTTAAACTTCATCTCTTTGGCCACT 316	
Db	287	LysThrIleAspIleThrAspLeuLeuAspSerIleProMetAsnGluGlnLeuLysVal 305	
QY	317	GTGAAAAGCAACTG-----AGTAGAGTCANATCAAGAAACACAGCTGGAGTTA 364	
Db	307	SerLeuLysProLeuValSerGluSerSerSerSerLys-----LysLeuAspAla 325	
QY	365	CTCTGTAACAAAGAGATGTAACGGATCCACAGAGAA---TAGCATTCATAAAGACGC 421	
Db	326	ProLeuAlaLysSerIleGlnAspArgLeuGluArgGlnAlaAlaTyrGluGlnThrLys 345	
QY	422	ACAAGTCTCTCCAAATGGGACCTGCTCCTGCTGAAGACCGGACGACGAGCTGGT 481	
Db	346	AsnAsp-LeuGluLysTrpLysProIleValAlaAspAsnArgLysSerAspGluLeuI 365	
QY	482	TTTTCCCTCGGAAAGAGGAGGAGCCATGCTGCCATTTGAACATGCTGCTAGTGGCTG 541	
Db	365	ePheProMetAsnGluThrAlaArgProValProSerAsnAsnGlyLeuAlaSerSerPh 385	
QY	542	GAGGCAAGAACTCCCTCGGACGAGGAGAAATTTTCAAC---CTCCTCCATTAAGAACAGCA 598	
Db	385	eGluProArgThrGluSerGluArgLysMetHisGlnAlaLeuLeuAspAlaGlyLeuG 405	
QY	599	GCACGTGACAGACCTTTACTGACCCCTGTGGAAAAGCCCTCTCTCGGACCCATGAGCCT 658	
Db	405	u-----AsnGluSerAlaLeuLysLysGlnGluGluLeuAlaLeuAsnLysLeuSerVa 423	
QY	659	AGAAGAGCAAGATGCGACGACGAGCTTCAGAGGGCTCGGGCTCTGCGCTCTACTACTA 718	
Db	423	iGluGluValAlaGluArgThrArgGlnLeuArgPheMetArgGluLeuMetPheArgG 443	
QY	719	TGAGCCCAAGGCTCGAAGACGAGAGAAATCTNAAAGTATAAGTATACACAAAGTCGTGA 778	
Db	443	uGluArgLysAlaLysArgValAlaLysIleLysSerLysThrTyrArgLysIleArgLy 463	

QY	779	GAAAGGAAGGCAAGAAA---GCCCTA-----AAAGAGTTTACAGCAGCTCGGAAGGT 829	
Db	463	sAsnArgLysGluLysGluMetAlaLeuIleProLysSerGluGluAspLeuGlu----- 481	
QY	830	TATCCAGCTGCCGCTACTAGAACACGACAAAGAGGAGGAGGAGGAGGAGGAGGAGAA 889	
Db	482	-AsnGluArgIleLysSerGluGluAlaArgAlaLeuGluArgMetThrGlnArgHisLy 501	
QY	890	GAAG----- 893	
Db	501	sAsnThrSerSerTrpThrArgLysMetLeuGluArgAlaSerHisGlyGluClyThrAr 521	
QY	894	-----ACAAGGAGAGAGAAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 943	
Db	521	gGluAlaValAsnGluGlnIleArgLysGlyAspGluLeuMetGlnArgIleHisGlyLy 541	
QY	944	AGAGGAAAGGAA---GAAGGAGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 992	
Db	541	sGluIleSerGluMetAspGlyGluAspValSerGluPheSerAspSerAspTyrAspTh 561	
QY	993	-GGAGAGAAAGAAAGAACTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAA 1051	
Db	561	rAsnGluGlnValSerThrAlaPheGluLysIleArgAsnGluGluGluProLysLeuLy 581	
QY	1052	AAAGTN-----AAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1082	
Db	581	sGlyValLeuGlyMetLysPheMetArgAspAlaSerAsnArgGlnLysAlaLeuValGl 601	
QY	1082	----- 1082	
Db	601	nAspGluMetGlnAlaPheGluAspGluLeuAlaGlyValProAsnGluAspThrSe 621	
QY	1083	-----GAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1102	
Db	621	rGlnLysGlyGluAspGlyValProGlyValLeuIleGlyAsnAsnThrGlyArgArgse 641	
QY	1103	GAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1162	
Db	641	rPheLysProSerGluGluAlaAlaLysLeuSerLeuProSerArgLysAsnProPheVa 661	
QY	1163	GAATAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1222	
Db	661	lSerAspSerAlaValLeuLysValAsnLysProGluMetLysGluGlyCln----- 678	
QY	1223	AAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1256	
Db	679	---LysLysAlaGluAlaArgLysLysLysGlu 688	

RESULT 14
NFM_MOUSE
ID NFM_MOUSE
AC P08553; Q61961;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
OS (Neurofilament medium polypeptide) (NF-M).
GN NEF3 OR NEFM OR NFM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87246694; PubMed=3036526;
RA Levy E., Liem R.K.H., D'Eustachio P., Cowan N.J.;
RT "Structure and evolutionary origin of the gene encoding mouse NF-M,
the middle-molecular-mass neurofilament protein.";
RL Eur. J. Biochem. 166:71-77(1987).
RN [2]
RP SEQUENCE OF 322-540 FROM N.A.
RX MEDLINE=87158637; PubMed=3103856;
RA Julien J.-P., Meyer D., Flavell D., Hurst J., Grosveld F.;

D	b	612	lGluValLysProLysProGluAlaLysAlaGlyLysGlyGluInLysGluGlUGI	632
Q	y	1121	AAGAAGAAAGAAATAAGGAACNAGA -AAGAAGGAGAAGAAAGAAATTAAGAGAGAACA	1179
D	b	632	UlysValGluGluGluLysLysGluValThrLysGluSerProLysGluLysValGI	652
Q	y	1180	AGAAAAGAAAGAAAGAAAGGAAGGAAGGAAAAAGGAAAAAGGAAGGAAGAAA	1239
D	b	652	uLysLysGluGluLysProLysAspValalaAsplyLysLysAlaGluSerProValY	672
Q	y	1240	GTAGAAACGGGAAGAACA-----	1258
D	b	672	sGLuLysAlaValGluGluValIIeThrLysSerVallLysValSerLeuGluY	692
Q	y	1259	-----AAAGAAAGTATAAGAAAGGAAGAAAGAAAGAACA	1290
D	b	692	sASPThrLysGluGluLysProGlnProGlnGluLysVallLysGluLysAlaGluGlUGI	712
Q	y	1291	AGGA 1294	
D	b	712	ugly 713	
 RESULT 15 ATTR_CABEL				
ID	ATTR_CABEL	STANDARD;	PRT; 1359 AA.	
AC	Q9U7EO; 002061;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).			
GN	XNP-1 OR B0041.7.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;			
OC	Rhabditiidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.			
RX	MEDLINE=99365296; PubMed=10433961;			
RA	Villard L., Fontes M., Ewbank J.J.;			
RT	"Characterization of xnp-1, a Caenorhabditis elegans gene similar to			
RT	the human XNP/TRP-X gene";			
RL	Gene 236:13-19(1999).			
RZ	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-Bristol N2;			
RA	Fulton R., Wohldmann P.;			
RU	Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.			
CC	-I- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES			
CC	GENE EXPRESSION BY AFFECTING CHROMATIN (Potential).			
CC	-I- SUBCELLULAR LOCATION: Nuclear.			
CC	-I- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; AF134186; AAD55361.1; .			
DR	EMBL; AF000196; AAC24256.1; .			
DR	WormPeP; B0041.7; CE17314.			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR001650; Helicase_C.			
DR	InterPro; IPR000330; SNF2_N.			
DR	Pfam; PF00176; SNF2_N; 1.			
DR	Pfam; PF00271; helicasC; 1.			
DR	SMART; SM00487; DEXDC; 1.			
DR	SMART; SM00490; HELICc; 1.			
KW	DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.			
FT	NP_BIND 496 503			
FT	SITE 636 639			
FT	DEAH BOX			

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:20:42 ; Search time 44.5135 Seconds

(without alignments)

12016.549 Million cell updates/sec

Title: US-09-502-945-3

Perfect score: 2140

Sequence: 1 ggcgtgtaactgactgcga.....gaagaagaagaaggaagaaaaa 1298

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spo1/US09502945/runat_14032003_101058_19113/app_query.fasta_1.10979
-DB-SPTREMBL_21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOEPCI=0 -LOEPCXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09502945.ecgn_1.1.565.ernat_14032003_101058_19113 -NCPG=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organella:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.rvirus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1338.5	62.5	763	4 Q8WWS6	Q8WWS6 homo sapien

2	1338.5	62.5	771	4 Q9BVJ6	Q9BVJ6 homo sapien
3	1183	55.3	766	4 Q92555	Q92555 homo sapien
4	600	28.0	131	4 O60531	O60531 homo sapien
5	571	26.7	194	11 Q9CZK8	Q9CZK8 mus musculus
6	403	18.8	260	10 Q9LGG9	Q9LGG9 arabidopsis
7	365	17.1	554	5 Q9VPS3	Q9VPS3 drosophila
8	322.5	15.1	699	10 Q9ZU69	Q9ZU69 arabidopsis
9	312.5	14.6	1701	5 O61164	O61164 plasmodium
10	312	14.6	2274	5 Q9VYU0	Q9VYU0 drosophila
11	310.5	14.5	685	10 Q8W3X8	Q8W3X8 cucurbita m
12	309.5	14.5	771	5 Q9VUR2	Q9VUR2 drosophila
13	309.5	14.5	2055	5 Q8F5C7	Q8F5C7 plasmodium
14	302	14.1	2081	10 Q9LH98	Q9LH98 arabidopsis
15	299	14.0	695	5 Q9BHM3	Q9BHM3 paramecium
16	287.5	13.4	335	5 O76719	O76719 caenorhabdi
17	285	13.3	1166	10 Q9SYP6	Q9SYP6 arabidopsis
18	281	13.1	735	4 Q9UEQ5	Q9UEQ5 homo sapien
19	281	13.1	784	4 Q9UIE9	Q9UIE9 homo sapien
20	280	13.1	385	5 Q17909	Q17909 caenorhabdi
21	280	13.1	460	2 Q9ZIU2	Q9ZIU2 borrelia bu
22	279	13.0	369	16 Q83483	Q83483 treponema p
23	279	13.0	876	11 Q9WU62	Q9WU62 mus musculus
24	278.5	13.0	499	10 Q9FYW3	Q9FYW3 lycopersico
25	278	13.0	1108	5 Q9NDI0	Q9NDI0 babesia big
26	277	12.9	513	10 Q9LW95	Q9LW95 nicotiana t
27	276	12.9	1339	11 Q35788	Q35788 rattus norv
28	275	12.9	451	16 Q50870	Q50870 borrelia bu
29	274.5	12.8	390	5 Q18401	Q18401 caenorhabdi
30	274	12.8	1062	5 Q960C4	Q960C4 drosophila
31	272	12.7	647	5 P91280	P91280 caenorhabdi
32	272	12.7	919	4 Q9NQS7	Q9NQS7 homo sapien
33	269.5	12.6	494	11 Q9JKL7	Q9JKL7 rattus norv
34	268.5	12.5	1750	3 Q8XOH2	Q8XOH2 neospora
35	266	12.4	508	4 Q8WXA9	Q8WXA9 homo sapien
36	266	12.4	1052	5 Q26938	Q26938 trypanosoma
37	264.5	12.4	268	6 Q9SL36	Q9SL36 tryctolagus
38	264.5	12.4	312	5 P91570	P91570 caenorhabdi
39	264.5	12.4	810	10 Q9ZW13	Q9ZW13 cucurbita m
40	264	12.3	1430	5 Q9W0B0	Q9W0B0 drosophila
41	259	12.1	1432	10 Q23230	Q23230 arabidopsis
42	257.5	12.0	901	5 Q8SUW1	Q8SUW1 encephalito
43	257	12.0	546	4 Q9NW40	Q9NW40 homo sapien
44	256.5	12.0	1008	10 Q9FNE4	Q9FNE4 arabidopsis
45	256	12.0	873	13 O13024	O13024 xenopus lae

ALIGNMENTS

RESULT 1

ID	Q8WWS6	PRELIMINARY;	PRT;	763 AA.
AC	Q8WWS6;			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DE	DJ537K23.3 (KIAA0266 protein) (Fragment).			
GN	DJ537K23.3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RA	SEQUENCE FROM N.A.			
RP	Grafham D.			
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL034405; CAD20048.1; -			
FT	NON_TER 763			
SQ	SEQUENCE 763 AA; 87075 MW; 41E3979FF06825C9 CRC64;			

Alignment Scores:

Pred. No.:	5.74e-85	Length:	763
Score:	1338.50	Matches:	323
Percent Similarity:	63.62%	Conservative:	32

```

Best Local Similarity: 57.89%      Mismatches: 72
Query Match: 62.55%      Indels: 133
DB: 4      Gaps: 8

US-09-502-945-3 (1-1298) x Q8WWS6 (1-763)

```

Qy	11	ATGACTCGCAACCGGCTTGCGAGAGACCTTCCTGGCTTTGAGCCANCAAGAGAACTACGC	70
Db	1	MetThrAlaasnArgLeuAlaGluSerLeuLeuAlaLeuSerGlnGlnGluLeuAla	20
Qy	71	GATTTCGCAAAAGACTACCTTTGAGTGAGAGTGAAGATGAGGGGACAATGATGGAGAG	130
Db	21	AspLeuProLysAspTyrLeuLeuSerGluSerGluAspGluGlyAspAsnAspGlyGlu	40
Qy	131	AGAAAGCATCNAAAGCTTCTGGGAAGCAATCAGTTCCTCTTGATGGAAAGTAAGCGGAAA	190
Db	41	ArgLysHisGlnLysLeuLeuGluAlaIleSerSerLeuAspGlyLysAsnArgArgLys	60
Qy	191	TTTGCTGANAGGCTGAGGCTACTCGAAGGTCTCAGAGTTCAAATCAGTTCCTCAAGGA	250
Db	61	LeuAlaGluArgSerGluAlaSerLeuLysValSerGluPheAsnValSerSerGluGly	80
Qy	251	TCAGGAGAAAAGCTGGTCTTGCGAGATCTGCTTGAGCCCTGTTAAACCTTCATCTTCTTG	310
Db	81	SerGlyGluLysLeuValLeuAlaAspLeuLeuGluProValLysThrSerSerSerLeu	100
Qy	311	GCCACTGTGAAAAGCAACTGAGTAGAGTCANATCAAAGAAANCAGTGGAGTACTCTGT	370
Db	101	AlaThrValLysLysGlnLeuSerArgValLysSerLysLysThrValGluLeuProLeu	120
Qy	371	AACAAAGAAGAGATTGAACGGATCCACAGAGATA- GCATTCAATAAAACGCA-CAAGTC	428
Db	121	AsnLysGluGluIleGluValGluIleHisArgLysValAlaPheAsnLysThrAlaGlnVal	140
Qy	429	CTCTCCAAATGGGACCTGTCTCGTGAAGAACCGCAGCAGCAGAGCAGCTGGTTTTCCC	488
Db	141	LeuSerLysTyrAspProValValLeuLysAsnArgGlnAlaGluGlnLeuValPhePro	160
Qy	489	CTGGAGAAGAGGAGCCAGCATTCCTCCCATTTGACATCTGCTCAGTCGGCTGGGAAGCA	548
Db	161	LeuGluLysGluGluProAlaIleAlaProIleGluHisValLeuSerGlyTyrLysAla	180
Qy	549	AGAACTGCTCCCTGGACGAGGAAATTTCAACCTCTCTCATAGAACAAGCAGCCAGTGACA	608
Db	181	ArgThrProLeuGluGlnGluIlePheAsnLeuLeuHisLysAsnLysGlnProValThr	200
Qy	609	GACCTTTACTGACCCCTGTGGAAAAGGCTCTCTCCGAGCCATGAGCCTAGAAGAGCA	668
Db	201	AspProLeuLeuThrProValGluLysAlaSerLeuArgAlaMetSerLeuGluGluAla	220
Qy	669	AAGATGCCACAGCAGAGCTTCAGAGGGCTCGGGCTCTGCAGTCTTATGAGCCCAAG	728
Db	221	LysMetArgArgAlaGluLeuGlnArgAlaArgAlaLeuGlnSerTyrTyrGluAlaLys	240
Qy	729	GCTCGAAGAGAGAAGAAATTCNAAAGTTAAAGTATTCACAAAGTCGTGAAGAAAGGAAG	788
Db	241	AlaArgArgGluLysLysIleLysSerLysLysTyrHisLysValValLysLysGlyLys	260
Qy	789	GCCAGAAGCCCTAAAGAGTTTGACAGCTCGGGAAGGTTAATCCAGCTGCCGCACCTA	848
Db	261	AlaLysLysAlaLeuLysGluPheGluGlnLeuArgLysValAsnProAlaAlaLeu	280
Qy	849	GNAGAACAAGAAAGAGGAAGAGGAGGAGGAGGAGGAAG- - - - -	890
Db	281	GluGluLeuGluLysIleGluLysAlaArgMetMetGluArgMetSerLeuLysHisGln	300
Qy	891	-----AAGAAACAAG-----	899
Db	301	AsnSerGlyLysTyrPalaLysSerLysAlaIleMetAlaLysTyrAspLeuGluAlaArg	320
Qy	900	-----GAGAAGAAGAAAGAGGGGAGAGGAGAGAGAAAGAAAGAA	938
Db	321	GlnAlaMetGlnGluGlnLeuSerLysAsnLysGluLeuThrGlnLysLeuGlnAla	340

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojofori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fledrichmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,


```
QY 689 TCAGAGGGCTCGGGCTCTCGAGTCTTACTATGANGCCAAAGGCTGAAGAGAGAGAAAT 748
    |||
Db 488 GluGlu-----GluAlaLysArgArgGluGluGlu 497
QY 749 CNAAGTTAAAGTATCAAAAGTCTGTGAAGAAAGGCAAGAGCCCTAAAGA 808
    |||
Db 498 ArgLysArgGluGluGluAlaGluGlnAlaArgLysArgGluGluGluArgGluLys 517
QY 809 GTTTGACGAGCTCGGAAGTTAATCCAGCTCGCAGCTAGAAACCAAGAAAGAGGA 868
    |||
Db 518 Glu-GluGluMetAlaLys-----LysArgGluGluGlu 528
QY 869 AGAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 928
    |||
Db 528 uArgGlnArgLysGluGluGluValGluArgLysArgGluGluGluGluGluArgLys 548
QY 929 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 979
    |||
Db 548 sArgArgGluGluGluAlaArgLysArgGluGluGluArgLysArgGluGluMetAl 568
QY 980 AAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1033
    |||
Db 568 aLysArgArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 588
QY 1034 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075
    |||
Db 588 gGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 608
QY 1076 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1135
    |||
Db 608 nLysLysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 628
QY 1136 TAAGGACACNAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1186
    |||
Db 628 uGluGluMetAlaLysLysLysLysLysLysLysLysLysLysLysLysLys 648
QY 1187 CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
    |||
Db 648 uArgLysArgArgGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 668
QY 1238 AACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
    |||
Db 668 uAlaAlaLysArgAlaGluGluGluGluGluGluGluGluGluGluGluGlu 687

RESULT 9
O61164 PRELIMINARY; PRT; 1701 AA.
AC O61164;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Erythrocyte binding protein.
GN MAEBL.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM.
RX MEDLINE=98115903; PubMed=9448314;
RA Kappe S.H.I., Noe A.R., Fraser T.S., Blair P.L., Adams J.H.;
RT "A family of chimeric erythrocyte binding proteins of malaria
   parasites."
RL Proc. Natl. Acad. Sci. U.S.A. 95:1230-1235(1998).
DR EMBL: AF031886; AAC05366.1;
SQ SEQUENCE 1701 AA; 199268 MW; EDA8E2DEFD87CE8A CRC64;
```

Alignment Scores:

```
Pred. No.: 6.72e-14 Length: 1701
Score: 312.50 Matches: 99
Percent Similarity: 52.17% Conservative: 69
Best Local Similarity: 30.75% Mismatches: 114
```

```
Query Match: 14.60% Indels: 40
DB: 5 Gaps: 12
US-09-502-945-3 (1-1298) x O61164 (1-1701)
QY 417 AACGCACAGTCCCTCCAAA-----TGGGACCCCTGTCGTC 452
    |||
Db 1098 AsnAlaGluIleIleArgLysPheGluGluAlaGlnLysAlaIleAlaLysLysAla 1117
QY 453 CTGAAGAACCCGCGCAGCAGCAGCTGTTTCCCTTGGAGAAAGAGAGCCAGCCATT 512
    |||
Db 1118 GluGluGluArgLysLysAlaValLysLysAlaGluGluGluArgLysArgLys 1137
QY 513 -----GCTCCCATTTGAACATGTCGTAGTGGCTCGAAGGCAAGCAACTCCCTCGAG 563
    |||
Db 1138 GluAlaGluLysLysAlaGluGluArgLysArgLysArgLysAlaGluLysLysAlaGlu 1157
QY 564 CAGGAAATTTTCACCTCCTCCATTAAGAACAGCAGCCAGTGCACAGACCCCTTACTGACC 623
    |||
Db 1158 GluGluArgLysArgLysAlaGluGluLysLysAlaGluGluGluArgLysLysAlaGlu 1177
QY 624 CCGTGTGGAAGGCGCTCTCTCCGAGCCATGAGCTAGAGAGGCAAGATGCGAGCAGCA 683
    |||
Db 1178 AlaAlaLysLysAlaGluGluGluArgLysArgLysAlaGluGluLysLysAlaGluGlu 1197
QY 684 GAGCTTCAGAGGCTCGGGCTCTCGAGTCTTACTATGAGCCAGGCTCGAGAGAGAG 743
    |||
Db 1198 GluArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1212
QY 744 AAAATCNAAGTTAAAGCTATCACAAAGTCTGTGAAGAAAGGAGGCAAGAA----- 797
    |||
Db 1213 Lys-----LysAlaGluAlaValLysLysLysAlaGluGluAlaLysLysLysAla 1228
QY 798 ---GCCCTAAAGAGATTTGAGCAGCTCGGAGAGGTTAATCCAGCTGCC-----ACACTA 848
    |||
Db 1229 GluAlaAlaLysLysAlaGluGluGluArgLysLysLysAlaGluAlaLysLysAlaLeu 1248
QY 849 GAAGAACGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 908
    |||
Db 1249 GluArgLysLysLysSerGlu-----AlaAlaLysLysAlaLeuGluGluArgLysLys 1266
QY 909 AAAGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 968
    |||
Db 1267 AlaGluAlaAlaLysLys-AlaGluGluLysLysLysLysAlaGluAlaAlaLysLysAlaGlu 1286
QY 969 AAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1028
    |||
Db 1286 uGluGluLysLysLysAlaGluAlaLysLysAlaGluGluGluLysLysArgLysLys 1306
QY 1029 AATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1085
    |||
Db 1306 sLeuGlnArgLysProGluArgLysLysLysLysLysLysLysLysLysLysLysLys 1326
QY 1086 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137
    |||
Db 1326 gLysLysLysArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1346
QY 1138 ----AGGACNAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1193
    |||
Db 1346 gGlnArgLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1366
QY 1194 AGAAGAA-----GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1232
    |||
Db 1366 sGluLysLysAlaGluAlaAlaLysLysLysAlaGluGluGluArgLysLysLysLys 1386
QY 1233 GAAGAAAGTAGAAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1292
    |||
Db 1386 uLysLysAlaGluGluGluGluArgLysLysAlaGluAlaAlaLysLysLysLysLys 1406
QY 1293 GAAA 1296
    |||
Db 1406 gLys 1407
```

RESULT 10

Q9YVU0 PRELIMINARY; PRT; 2274 AA.

AC Q9YVU0; 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)

DE CG1905 protein.

GN CG1905.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Balwle R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleeb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).

DR EMBL; AE003487; AAF48098.1;

DR FlyBase; FBgn0030337; CG1905.

SQ SEQUENCE 2274 AA; 247009 MW; B287311CBDF9BF20 CRC64;

Alignment Scores:

Pred. No.: 7,17e-14 Length: 2274

Score: 312.00 Matches: 128

Percent Similarity: 47.31% Conservative: 83

Best Local Similarity: 28.70% Mismatches: 136

Query Match: 14.58% Indels: 99

DB: 5 Gaps: 16

US-09-502-945-3 (1-1298) x Q9YVU0 (1-2274)

QY 101 AGTGAAGATGAGGGGACAAATGATGGAGAGAGAAAGCATCNAAGCTTCTGGAAGCAATC 160

|||||

Db 1203 SerGluaspGluAlaThr-----ThrThrIleThrLeuAlaLys 1215

QY 161 AGTTCCCTTGATGGAAAGAAATAGCGGAAATTTGGCTCGANAGGTCTGAGGCTAGTCTCTCAAG 220

|||||

Db 1216 SerAlaLysLysAlaLysAsnAsnLysGln-----LysThrSerGlyLys 1230

QY 221 GTGTCAGAGTTCAATGTCAGTTCTGAAGGATCAGGAGAAAAGCTGGCTCTTCGACAGATCTG 280

|||||

Db 1231 ValSerGluLysLysProGluGlnThrAsnLeuAlaGluGluLysVal----- 1246

QY 281 CTTGAGCCTGTAAACCTTCTCTTGGCCACTCTGAAAAGCAACTGAGTAGATC 340

|||||

Db 1247 -----AlaValGluLysGluGluAsnGluGlu 1255

QY 341 ANATCAAGAANACAGTGGAGTTACCTCTGAACAAAGAA-----GAGATTGAACGG 391

|||||

Db 1256 ArgAlaValProMetValLysLysThrIleAsnLysGluGluAspSerAspAlaGluSer 1275

QY 392 ATCCACAGAGAAATAGCATTCATAAAGCGCAGCAAGCTCTCCAAATGGGACCTGTCGT 451

|||||

Db 1276 AsphIsalaAsp-SerLeuLeuAlaAsnLysSerSerIleAla-----AlaValMe 1292

QY 452 CCTGAAGAACCGCGCAGCAGCAGCGCTGTTTCCCTCGAGAAAGAGAGCCGCCAT 511

|||||

Db 1292 tValSerSerAlaSerAlaGlnGlyLeuSerLeuHisValGlu----- 1306

QY 512 TGTCCCATTTGAACATGTGCTCAGTGGCTGGAAGGCAAGAACTCCCTGGAGCAGGAAAT 571

|||||

Db 1307 -----MetSerAlaAlaAspAlaGluGlnGlyGluaspGluLul 1320

QY 572 TTTCAACTC-----CTCCATAAGAACACAGCAGCC 601

|||||

Db 1320 eGluGlyLeuAspGluGluProLysThrMetSerLysAspAsnLysLysLysGlnLy 1340

QY 602 AGTGACAGACCTTTACTGACCCCTGTG-----GAAAGCCCTCTCTCCG 646

|||||

Db 1340 sProGlyAspAlaValAlaThrMetThrIleAspLysGluLysGluLysAlaLysGluLy 1360

QY 647 AGCCATGAGCTAGAGAGCAAGAGTGCAGCAGCAGCAGCTTCAG----- 692

|||||

Db 1360 sGluLeuLysLeuLysGlu---LysGluArgGluAlaLysLeuGlnGluLysGluLysG 1379

QY 693 -----AGGCTCGGCTCT 706

Db 1379 uGluLysLeuLysLeuLysGluArgGluGluSerLeuArgMetGluArgGluGluLysLe 1399

QY 707 GCAGTCTCTACTATGANGCCAGGCTCAAGAGAGAGAGAAATCNAAGTTTAAAGTATCA 766

|||||

Db 1399 uLysGluGluLysIleLysGluLysGluArgGluGluLysLeu-----LysGluG 1416

QY 767 CAAGTCGTGAAGAAAGGAGCCCAAGAGCCCTTAAAGAGTTTGACGAGCTCGGAA 826

|||||

Db 1416 uLysIleLysGluLysGlnArgGluGluLys---LeuLysGlu---GluLysLeuLysG 1434

QY 827 GGTATTCAGCTGCCCTCCTAGAACACCAAGAGAGGAAGAA-----AGGAG 877

|||||

Db 1434 uLysGluArgGluGluArgMetLysGluLysGluArgGluGluLysAlaLysGluLysG 1454

QY 878 GAGGAGAAAGCAAGCAAG 937

|||||

Db 1454 nArgGluGluLysLeuArgGluGluLysIleLysGluLysGluArgGluGluLysLeuL 1474

QY 938 AGGAG 985

|||||

Db 1474 ysGluLysLeuArgGluGluLysIleLysGluLysGluLysGluLysGluLysGluLysG 1494

QY 986 AAAAG 1045

|||||

Db 1494 luArgGluGluLysMetArgGlu---LysGluArgGluGluLysIleLysGluLysGluAr 1513

QY 1046 AAAG 1105

|||||

Db 1513 gValGluLysIleLysGluLysGluArgGluGluLysLeuLysLysGluGluLysGluLy 1533

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-F., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB003530; AAF49613.2;
 DR EMBL: AY069576; AAL39721.1;
 DR Flybase: FBgn0036514; CG12301.
 DR SEQUENCE 771 AA; 88727 MW; 912B33FE58D7CEDD CRC64;

Alignment Scores:

Pred. No.: 1,13e-13 Length: 771
 Score: 309.50 Matches: 123
 Percent Similarity: 45.15% Conservative: 77
 Best Local Similarity: 27.77% Mismatches: 180
 Query Match: 14.46% Indels: 63
 DB: 5 Caps: 15

US-09-502-945-3 (1-1298) x Q9VUR2 (1-771)

QY 101 AGTGAAGTGGGGGACAAATGATGGAGAGAGAAAGCATCAAGCTCTCGAAGCAATC 160
 DB 1 SerAspAspGlu---GluHisTyrGluProLysAlaHisLysLysLeuLeuGlnAlaIle 20
 QY 161 AGTTCCTTCATGGAAGAATAGCGGAATGGCTGACAGAGTCTGAGGCTAGTCTGAAG 220
 DB 21 GlySerGluGlyLysValGlnHisIleGlnLysSerThrArgAspGluArgGlnSerLeu 40
 QY 221 GTCTCAGAGTCTC-----AATGTCAGTCTCAAGCATCAGAGAA 259
 DB 41 GlnAspGluPheGlnLeuValLysGlyValSerSerAlaGluThrAspHisAlaProArg 60
 QY 260 AACTGTCTTCGAGATCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 319
 DB 61 AlaValGlyLeuAsnAspLeuValAspIleLeuArgThrSerThrLysHisSerGlnThr 80
 QY 320 AAAAAGCACTGAGTAGAGTC---ANATCAAGAAACAGTGGAGTACCTCTGACACAA 376
 DB 81 GlyLysLysLysLysAsnIleHisGlySerLysValLysValLysValLysValLys 100
 QY 377 GAAGAGATTGAACGGATCCACAGAGATAGCATCAATAAAGCAAGCAAGTC----- 428
 DB 101 ProAlaAlaAspArg-----IleLysArgThr-IleGlyTyrGln 113
 QY 429 -----CTCTCAAAAGGGACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 472
 DB 113 uGlyValThrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 133
 QY 473 GCAGCTGTTTTCCTCGAGAGAAAGAGGAGCCAGCATTCCTCCCATTCGACATGTGCT 532
 DB 133 uThrGlnIlePheProLeuProSerGluThrValTyrValAsnThrAlaAlaAsnAlaAr 153
 QY 533 CAGT---GGCTGGAAGGCAAGAACTCCCTGGAGCAGGAAATTT-----TTCAA 577
 DB 153 gProLeuAsnThrArgValLysSerAsnLeuAlaGlnGluLeuGluAlaAsnAsnArgLy 173
 QY 578 CTTCTCCTCAATAGAACAGCAGCAGTGTGACAGCCCTTTA-----CTGACCCC 625
 DB 173 sLeuArgGluLeuArgLysAlaGlnIleGlyAspThrThrAspGluLysGluLeuAlaLy 193

QY 626 TGTGGAAAAGGCGCTCTCTC---CGAGCCATGACCTAGAGAGGCAAGATCGACGAGC 682
 DB 193 sGlnGluArgLeuLeuGlnLysLysLysLeuThrArgAspGluLeuPheAlaArgArgLy 213
 QY 683 AGAGCTTCAGAGGCTCGGGCTCTGCAGTCTCTACTATGANGCCCAAGCTCGAAGAGAA 742
 DB 213 sGluLeuAlaTyrLeuLysMetArgGluSerGlnLysSerAlaLysAlaArgMetGlnAs 233
 QY 743 GAAATCNAAAGTTAAAGTATCACAAGTCGTGAAGAAAGGAAAGCCCAAGAAAGCCCT 802
 DB 233 nLysIleLysSerLysLysPheHisLysLysLeuGlnLysArgGlnLysMetLeuGlnMe 253
 QY 803 AAAAGAGTTTTCAGCTCGGAGCTTATCCAGCTGCC-----GC 844
 DB 253 tLysGluPheGluLeuGlnLysThrAspProGluAlaAlaLeuGluLysLeuAsnAl 273
 QY 845 ACTAGAGAACGAGAAAG 904
 DB 273 aLeuGluLysSerArgValGlnGluArgAlaSerLeuArgHisLysAsnThrGlyThrTr 293
 QY 905 GAAGAAAG 964
 DB 293 pAlaLysAsnLeuGlnIleArgAlaLysTyrAspLysAspValArgLysAspLeuAlaGln 313
 QY 965 AGAAAG 1024
 DB 313 uGlnLeuAlaValSerArg-----GluLeuThrGlnLysLysGlnValSe 328
 QY 1025 GAAGAAAG 1078
 DB 328 rAspSerGluAspGluThrLysArgValLeuProGluGluGluAspAsnAspTy 348
 QY 1079 GAAGAAAG 1138
 DB 348 rAspProPheAsnProThrLysArgLysAlaThrGluGlnThrGluAsnGlyAsnGln 368
 QY 1139 GGAACNAGAA-----AGAAG 1174
 DB 368 yGluSerGluGlySerAsnTrpArgGlnTyrTrpThrLysArgAsnLysAsnGluLysLe 388
 QY 1175 GAAGAAAG 1216
 DB 388 uLeuGlu-GluHisArgArgAspLeuGluGlyValGlnAspAsnAspGluGlnGluLul 408
 QY 1217 AGAAGAAAG 1276
 DB 408 ySgluThrIleArgLysProLeuIleGluValLysProAlaLysLysLysLysLys 428
 QY 1277 AAGAA 1281
 DB 428 erLys 429
 RESULT 13
 ID Q8T5C7 PRELIMINARY; PRT; 2055 AA.
 AC Q8T5C7;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE MaebI.
 GN MAEBL.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Blair B.L., Kappe S.H.I., Maciel J.E., Balu D.J., Adams B.;
 RT "Plasmodium falciparum maebI is a unique member of the ebl family."
 RL Mol. Biochem. Parasitol. 0:0-0(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Michon P., Stevens J.R., Kaneko O., Adams J.H.;
 RT "Plasmodium falciparum maebI is a unique member of the EBL family.";


```
QY 95 AGTGAGAGTGAAGATGAGGGGACAAATGATGAGAGAGAAAGCATCAATCAAGCTTCTCGAA 154
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 AsnLysLysGluSerGluLysValGluLysGlyLysLysGluSerLysAspAlaLys 797
QY 155 GCAATCAGTCCCTGTGATGGAAGAAATAGCGGAAATGGCTGANAGCTCTGAGGCTAGT 214
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 798 SerValGluThrLysAspAsnLysLysLysSerThrLysAsnArgAspGluAlaLys 817
QY 215 CTGAAGTGTCTCAGATTCATGTCAGTTCGAGGATCAGAGAAAGCTGGCTCTTGCA 274
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 818 GluArgSerGlyGluAspAsnLysGluAspLysGluSerLysAspTyrGlnSerVal 837
QY 275 GATCTGCTGAGCCTGTAAACATTCATCTCTTTGGCCACTG-----AAAACCAA 328
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 838 GluAlaLysGluLysAsnGluAsnGlyValAspThrAsnValGlyAsnLysGluAsp 857
QY 329 CTGAGTATAGATCANATCAAAANACAGTGGAGTTACTCTGACAAAGAGAGATTGAA 388
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 858 SerLysAspLeuLysAspArgSerValGluValLysAlaAsnLysGluGluSerMet 877
QY 389 CGGATCCACAGAGAATAGCATTAATAAAGCGACACAGTCTCTCCAAATGGACCTGT 448
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 878 LysLysLysArgGlu-----GluValGlnArgAsnAspLysSerThrLysGluVal 895
QY 449 CGTCTCT-----GAAGAACCGCGCAGCAGCAGCAGCTGGT 481
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 896 ArgAspPheAlaAsnAsnMetAspIleAspValGlnLysGlySerGlyGluSerValLys 915
QY 482 TTTTCCCTCGAGAGAGGAGCGACCATGCTGCCATTGCTCCCATGCTCAGTGGCTG 541
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 916 TyrLys----- 917
QY 542 GAAGGCAAGACTCCCTCGAGCAGGAGAAATTTCAACCTCTCTCATTAAGAACAGAGCC 601
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 918 -----LysAspGluLysLysGluLys-----LysGluGluAsn 929
QY 602 AGTGACAGACCTTTACTGACCCCTGTGGAAAGGCTCTCTCCGACCCATGAGCCTAGA 661
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 930 LysAspThrIleAsnThrSerSerLysGlnLysGlyLysAspLysLysLysLysLys 949
QY 662 AGAGGCAAGATCGGACGACGAGCTTCAGAGGCTCGGGCTCGACGCTCA----- 715
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 950 GluSerLysAsnSerAsnMetLysLysLysLysLysLysLysLysGluTyrValAsnAsn 969
QY 716 ---CTATGAGCCCAAGCTCGAAGACAGAGAGAAATCAATCAAGTAAAGTATCACAAGT 772
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 970 GluLeuLysLysGlnGluAspAsnLysLysGluThrLysSerGluAsnSerLysLeu 989
QY 773 CGTGAAGAAAGGAGCCCAAGAGCCCTAAAGAGCTTTGAGCAGCTCGGAGAGCTTAA 832
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 990 LysGluGluAsnLysAspAsnLysGluLysLys-----GluSerGlu 1003
QY 833 TCAGCTGCGCCACTAGACAGACAGAGAAAGAGGAGAAAGAGGAGGAGAGAGAAAGAA 892
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1004 AspSerAlaSerLysAsnArgLysLysLysGluTyrGluLysLysSerLysThrLys 1023
QY 893 GAACAAGGAGACAGACAGAAAGAGGAGACAGACAGAGAAAGAAAG----- 939
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1024 GluGluAlaLysLysGluLysLysLysSerGlnAspLysLysArgGluLysAspSer 1043
QY 940 GAGAAGAGAGAAAGGAGAGGAGAGAAAG-----AAGGAGAG 978
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1044 GluGluArgLysSerLysLysGluLysGluSerArgAspLeuLysAlaLysLysLys 1063
QY 979 GAAAGAGAAAGAGAGAGAGAAAGAGAA-----ACTAAGAGAGGAGAGG 1025
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1064 GluGluGluThrLysGluLysLysGluSerGluAsnHisLysSerLysLysLysGluAsp 1083
QY 1026 AGAATPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1085
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 LysLysGluHisGluAspAsnLysSerMetLysLysGluGluAspLysLysLysLysLys 1103
QY 1086 AGNAGAGGAGAGACTNAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1145
```

```
Db 1104 LysHisGluLysSerLysSerArgLysLysGluLysLysAspMetGluLysLeu 1123
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1146 GAA-----AGAAGGAGAGAGAAAGAAATAAGAGAGAGAGAGAGAGAGAGAG 1187
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1124 GluAspGlnAsnSerAsnLysLysLysGluAspLysAsnGlnLysLysSerGlnHis 1143
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1188 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1247
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1144 ValLysLeuValLysLysGluSerAspLysLysGluLysLysGluAsnGluLysSer 1163
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1248 CGG-----AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1262
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1164 GluThrLysGluLysSerLysSerLysSerGlnLysAsnGluValAspLysLysGluLys 1183
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1263 AAGATATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1184 LysSerLysAspGlnGlnLysLysLysGluLys 1195
```

RESULT 15

```
Q9BHM3
ID Q9BHM3 PRELIMINARY; PRT; 695 AA.
AC Q9BHM3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Cyclophilin-RNA interacting protein.
GN KIN241.
OS Paramecium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramecium.
OX NCBI_TaxID=5888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D4-2;
RA Krzywicka A.M., Beissson J., Keller A.M., Cohen J., Jerka-Dziadosz M.,
RA Klotz C.;
RT "KIN241 - a gene involved in cell morphogenesis in Paramecium
RT tetraurelia reveals a novel protein family of cyclophilin-RNA
RT interacting proteins (CRIPs) conserved from fission yeast to man.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D4-2;
RA Dessen P., Zagulski M., Gromadka R., Plattner H., Kissmehl R.,
RA Meyer E., Betermier M., Schultz J.E., Linder J.U., Pearman R.E.,
RA Kung C., Forney J., Satir B.H., Van Houten J.L., Keller A.M.,
RA Frolissard M., Sperling L., Cohen J.;
RT "Paramecium genome survey: a pilot project.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ409115; CAC35733.1; -
DR HSSP; P11940; ICVJ.
DR InterPro; IPR002130; CSA_PPIase.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00160; pro_isomerase; 1.
DR Pfam; PF00076; rrm; 1.
DR PRINTS; PR00153; CSAPPISMRASE.
DR PRINTS; PR01574; TUBBYPROTEIN.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50072; CSA_PPIASE_2; 1.
DR PROSITE; PS50102; RRM; 1.
SQ SEQUENCE 695 AA; 82646 MW; B8B583BFC5D50B89 CRC64;
```

```
Alignment Scores:
Pred. No.: 6,08e-13 Length: 695
Score: 299.00 Matches: 90
Percent Similarity: 51.86% Conservative: 91
Best Local Similarity: 25.79% Mismatches: 91
Query Match: 13.97% Indels: 77
DB: 5 Gaps: 10
```

US-09-502-945-3 (1-1298) x Q9BHM3 (1-695)

Search completed: March 21, 2003, 12:58:56
Job time : 64.5135 secs

```
QY 354 CAGTGGAGTTACTC-----TGAACAAGAAGAGAGATTGAACGGATCCACAGAG 401
||||| |||
Db 307 GlnTrpSerLeuLeuLysLysAlaGlnArgLysLysLysGluLeuAsnLeuGlnArgGlu 326
||||| |||
QY 401 ----- 401

Db 327 LeuAlaGlnArgSerLysAsnLysAspHisSerPheIleGluSerLeuLeuLysLys 346

QY 402 -----AATAGCATTCAATAAAAGCCACACAGTCCTCTCCAAA 437
||| |||
Db 347 AspTyrPheHisThrLysAlaAsnAspPheValIleAspGlnGlnLeuGlnLys 366
||||| |||
QY 438 TGGGACCTGTCTCTGTAAGAACCGCAGCAGCAGCTGGTTTTCCTCCCTGGAGAAA 497
||||| |||
Db 367 -----LeuGlnAspGlnGlnLeuLeuGlnMetGlnLeuAspSer 381
||||| |||
QY 498 GAGGAGCCACCATCTCCCATGTGAACATGTCTCAGTGGTGAAGCAAGACTCCC 557
||||| |||
Db 382 GlnGlnLysIle-----GlnGlnGlnThrGln 390

QY 558 CTGGAGCAGAGAAATTTTCAACCTCTCCATAAGACAAGCAGCCAGTCGACAGACCCCTTTA 617
||||| |||
Db 391 LeuLysGlnGlnLysLysLysLysLysLysLysAsnLysLys----- 404

QY 618 CTGACCCCTGTGGAAGAGCCTCTCTCCGAGCATGAGCCTAGAGAGGCAAGATGCCA 677
||||| |||
Db 405 ---LysSerAlaGlnLysLysAsnLysLysArgHisArgSerSerGlnGlnGluSerAsp 423
||||| |||
QY 678 CGAGCAGAGCTTCAGAGGCTCGGGCTCTGCTGCTCTACTATGANGCCCAAGGCTCGAAGA 737
||||| |||
Db 424 -----SerAspSerGlnSerLeuLysasn-----LysLysGlnLys 435

QY 738 GAGAAGAAATCAAAAGTTAAAGTATACAAAAGTCGTGAAGAAAGGAAGGCCAAGAAA 797
||||| |||
Db 436 AsplysHisLysAsnLysAsnLysHisLysSerTyrAspArgAspLysAspGlnGln 455
||||| |||
QY 798 GCCTTAAGACAGTTTGAGCAGCTGCGGAGGTTAATCCAGCTCCCTAGAGAACGA 857
||||| |||
Db 456 LysArgLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 468
||||| |||
QY 858 AGAAAGAGGAGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 917
||||| |||
Db 469 GlnGlnAspLeuGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 488
||||| |||
QY 918 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 977
||||| |||
Db 489 GluLysGluArg-GluLysGluLysGluLysGluLysGluLysGluLysGluLysGlu 508
||||| |||
QY 978 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1037
||||| |||
Db 508 gGluLysGlu---ArgGlnLysGluArgGlu---LysGluArgGluArgGluArgGluLys 526
||||| |||
QY 1038 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1097
||||| |||
Db 527 GluLysGluArgGluArgLysArgAspArgGluArgGluLysGluLysGluArgLysArg 546
||||| |||
QY 1098 ACTNAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1157
||||| |||
Db 547 GluLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 566
||||| |||
QY 1158 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1217
||||| |||
Db 567 LysArgGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 586
||||| |||
QY 1218 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1277
||||| |||
Db 587 GlyLysGlnGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 603
||||| |||
QY 1278 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298
||||| |||
Db 604 GlyLysAsnLysGlnSerGlu 610
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 08:38:27 ; Search time 54.6146 Seconds
(without alignments)
10910.938 Million cell updates/sec

Title: US-09-502-945-4

Perfect score: 4120

Sequence: 1 cctggccgcgtcggtcg.....tatttccagcttaaaaaa 2236

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O/cgn2_1/USPO1/US09502945/runat_14032003_101057_19087/app_query.fasta_1.10979
-DB-A_Geneseq_101002 -OEMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_365/runat_14032003_101057_19087 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	3371	81.8	652	20	AY07094
2	3371	81.8	652	20	AY07050
3	2331	56.6	481	20	AY07098
4	2311	56.1	521	20	AY07099
5	2046	49.7	403	20	AY07095
6	1732	42.0	328	20	AY07097
7	1350	32.8	268	20	AY38440
8	447	10.8	120	21	AA53328
9	425.5	10.3	493	22	AB861924
10	417	10.1	98	20	AY38390
11	406	9.9	95	20	AY38437
12	272	6.6	1037	22	ABG22366
13	271.5	6.6	928	23	AA211718
14	267.5	6.5	2037	21	AA533753
15	266.5	6.5	1111	23	AA087918
16	263.5	6.4	767	23	AA084269
17	263.5	6.4	767	23	AA084328
18	261.5	6.3	767	21	AA22138
19	261.5	6.3	1745	23	AB57253
20	261	6.3	1535	23	AA087934
21	259.5	6.3	856	21	AA01383
22	259.5	6.3	1881	20	AA24025
23	258.5	6.3	724	23	AB04798
24	258.5	6.3	1526	22	AA07977
25	256.5	6.2	1526	22	ABG06116
26	256.5	6.2	1526	22	ABG07289
27	256.5	6.2	1552	22	AA078793
28	256.5	6.2	1554	22	ABG07288
29	256.5	6.2	1584	22	ABG07290
30	256.5	6.2	1627	22	ABG06117
31	254.5	6.2	2070	20	AY04733
32	247.5	6.0	763	20	AY04741
33	247.5	6.0	1005	20	AY04731
34	247.5	6.0	1373	20	AY04730
35	247.5	6.0	2000	20	AY04732
36	242.5	5.9	724	23	AB04805
37	237	5.8	46	22	AB38777
38	237	5.8	46	22	AB23814
39	237	5.8	46	22	AA59416
40	237	5.8	46	22	AA071974
41	237	5.8	46	22	AA019386
42	237	5.8	46	22	AA032249
43	237	5.8	46	23	ABG41789
44	234.5	5.7	817	19	AA048101
45	234.5	5.7	817	20	AA030137

ALIGNMENTS

RESULT 1
AA07094
ID AA07094 standard; Protein: 652 AA.
XX
AC AA07094;
XX
DT 02-JUL-1999 (first entry)
XX
DE Colon cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.

XX PF 15-JUL-1998; 98WO-US14679.
 XX 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX PA Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 XX PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
 XX PI Tureci O;
 XX WPI: 1999-132448/11.
 XX New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 XX PS Disclosure; Page 656-658; 787pp; English.
 XX CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 XX SQ Sequence 652 AA;

Alignment Scores:
 Pred. No.: 2,75e-276 Length: 652
 Score: 3371.00 Matches: 652
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 81.82% Indels: 0
 DB: 20 Gaps: 0

US-09-502-945-4 (1-2236) x ANY07094 (1-652)

QY 97 ATGACCGAAAAGTGGCCGAGAAATCCGGCATAGGTGGATTTCTGATTGAAATGAT 156
 DB 1 MetAspArgLysValAlaArgGluPheHisLysValAspPheLeuLeuGluAsnAsp 20
 QY 157 GCAGAGAAGACATCTCTATGATGTCGCGAATGTACACACAGACCATGGACGTGGCC 216
 DB 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
 QY 217 GTGCTCGTGGAGACCTGAAGCTGGTCATCAATGAACCCAGCCGCTGCTCTGTTTGAT 276
 DB 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
 QY 277 GCATTCGGCCGCTGATCCACCTGAAGCACCAGGTGGATATGATCAGCTGACCCCGG 336
 DB 61 AlaIleArgProLeuLeuProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 QY 337 CGCTCAGGAAGCTGAAGAGGTGCGTCTGGACCGCTGACACCCGAGCCCTCGCCCTG 396
 DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
 QY 397 AGTGTGCTGGTGGCTGGATTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGGC 456
 DB 461 GluGlySerLeuAspPheLeuAlaLeuGluGlyValAspSerProIleGlyLysValVal 480

DB 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheHisSerHisLeuIleLysGly 120
 QY 457 GGTACAGCAGACAGCGTCCGGCTCCAGGTAGGGACGAGATCGTCCGATCAATGGATAT 516
 DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 QY 517 TCCATCTCTCTGTACCCATGAGAGGTGATCAACCTCATTCGACCAAGAAATCTGTG 576
 DB 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
 QY 577 TCCATCAAGTGAGACATCGGCTGATCCCGTGAAGAGCTCTCCTGATGAGCCCTC 636
 DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
 QY 637 ACTTGGCAGTATGTGGATCAGTTTGTGCGAATCTGGGGCGTGGAGGAGCCCTGGGC 696
 DB 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlySerLeuGly 200
 QY 697 TCCCTCGAATCGGAAACAGAGAGAGAGGTCTTCATCAGCCTGGTAGGCTCCGGA 756
 DB 201 SerProGlyAsnArgGluAsnLysLysValPheIleSerLeuValGlySerArg 220
 QY 757 GGCCTTGGCTGAGCATTTCCAGCGGCCCATCCAGAGCCCTGGCATCTTTATCAGCCAT 816
 DB 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
 QY 817 GFGAACTGGTCCCTCTGCTGAGTGGGATGGAGATAGGGGACAGATTCGTA 876
 DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
 QY 877 GTCATGCGCTGCTCTCTAACCTGATCACAGGAGGCTGTAATGTCTGAAAAAT 936
 DB 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
 QY 937 AGCCCGAGCCTGACCATCTCCATTTAGCTGAGCTGGCCGGAGCTGTTCATGACAGAC 996
 DB 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
 QY 997 CGGAGCGGCTGGCAGAGCGCGCTGAGCTGCGAGCGGAGGCTTCTCATGCGAG 1056
 DB 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuLeuMetGln 320
 QY 1057 AAGCGCTGGCGATGAGTCCCAACAGATCTCCAGGAGCAGCAGAGATGGAGCGCAA 1116
 DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGluGlnGlnGlnGlnGlnGln 340
 QY 1117 AGGAGAAAGAAATTCGCCAGAGGCAGCAGAGAGAAATAGAGATACCGAAGGAGATG 1176
 DB 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluAsnGluArgTyrArgLysGluMet 360
 QY 1177 GAACAGATTGAGAGAGAGAGAGAGTAAAGCAATGGGAAGAGACTGGGGCTCA 1236
 DB 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluLysTrpGlySer 380
 QY 1237 AAGGAACAGTACTCTGCTTAAACCATCACTGCTGAGGTACCCAGTACCCCTTCGC 1296
 DB 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 QY 1297 AAGCCAAAGTATGATCAGGAGTGAACCTGAGCTCGAGCCCGCAGATGACCTGGATGA 1356
 DB 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
 QY 1357 GGCAGGAGGAGCAGGAGAGAGGATTTCCGGAATATGAGGAAGGCTTTCACCCCTAC 1416
 DB 421 GlyThrGluGluGlnGlyGluGlnAspPheArgLysTyrGluGluGlyPheAspProTyr 440
 QY 1417 TCTATCTTCAACCCAGACAGATCATGGGAGGATGTCCGCTCTACCGATCAAGAG 1476
 DB 441 SerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLys 460
 QY 1477 GAGGATCTTACAGCTGGCCCTGGAAGCGGTGGAGTCCCCCATTTGGGAAGGTGTC 1536
 DB 461 GluGlySerLeuAspPheLeuAlaLeuGluGlyValAspSerProIleGlyLysValVal 480

QY 1537 GTTCTGCTGTATGAGCGGGAGCTGCTGAGCGGATGCGCATGTCGATTTGAAGGGGAC 1596
 Db |||||||
 481 ValSerAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAsp 500
 QY 1597 GAGATCATGCAATCAACGGCAAGATTGTGACAGACTACACCCCTGGCTGAGGCTGACGCT 1656
 Db |||||||
 501 GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAla 520
 QY 1657 GCCCTGAGAGGCTGGAATCAGGGCGGGGACTGAGTCGACCTTGCTGGTGGCTGCTGC 1716
 Db |||||||
 521 AlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValAlaValLys 540
 QY 1717 CCCCCAAGAGTATGACGATGACCTGACCTTCTGCTGAAGTCCAAAGGGGAAACAA 1776
 Db |||||||
 541 ProProLysGluTyrAspAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnGln 560
 QY 1777 ATTCACCGCTTAGGAACAGTACGCTCGGCCCCACCTCGTGACACAAAGCCTCGGACC 1836
 Db |||||||
 561 IleHisAlaLeuGlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgThr 580
 QY 1837 AGCCTTGAGAGAGCCACATGACACACACGATGGCATCTTGGACCTGAATCTATCA 1896
 Db |||||||
 581 SerLeuGluArgGlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSer 600
 QY 1897 CCCAGGAATCTCAACTCCCTTTGGCCCTGGAACAGGGCCAGATPAGGAACAGCTCGGC 1956
 Db |||||||
 601 ProArgAsnLeuLysLeuProLeuAlaLeuAsnGlnGlyGlnIleAlaGlnSerSerGly 620
 QY 1957 CACTTTTGAAGCCCAATGAGGAGGAAGGGAGCAGCCAGCCGTTTGGGAGAGATCTC 2016
 Db |||||||
 621 HisPheGluGlyGlnCysGlyGlyGlyAlaAlaSerArgLeuGlyGluAspLeu 640
 QY 2017 AAGGATCCAGACTCTCATCTCTTCTCTCTGCCCCAG 2052
 Db |||||||
 641 LysAspProAspSerHisSerPheProLeuAlaGln 652
 RESULT 2
 AAY07050.
 ID AAY07050 standard; Protein; 652 AA.
 XX AC AAY07050;
 XX DT 02-JUL-1999 (first entry)
 XX DE Renal cancer associated antigen precursor sequence.
 XX KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.
 XX OS Homo sapiens.
 XX PN WO9904265-A2.
 XX PD 28-JAN-1999.
 XX PF 15-JUL-1998; 98WO-US14679.
 XX PR 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX PA Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 XX WPI; 1999-132448/11.

XX

PT New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 for the diagnosis, monitoring or treatment of cancers
 PS Disclosure; Page 436-438; 787pp; English.
 XX
 CC The invention relates to a method for diagnosing a disorder characterised
 by expression of a human cancer associated antigen precursor coded for by
 a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 biological sample isolated from a subject with an agent that specifically
 binds to the NAM, an expression product or a fragment of an expression
 product complexed with an HLA molecule; and (b) determining the
 interaction between the agent and the NAM or the expression product as a
 determination of the disorder. The products and methods can be used in
 the diagnosis, monitoring, research, or treatment of conditions
 characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 which are cancer associated antigen precursors expressed in human breast
 cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 lung cancer.

SQ Sequence 652 AA;

Alignment Scores:

Pred. No.: 2,75e-276 Length: 652
 Score: 3371.00 Matches: 652
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 81.82% Indels: 0
 DB: 20 Gaps: 0

US-09-502-945-4 (1-2236) x AAY07050 (1-652)

QY 97 ATGGACCGAAAGTGGCCGAGAAATTCGGCATAGGTGGATTCTGTGATGAAATGAT 156

Db |||||||

1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20

QY 157 GCAGAGAAGGACTATCTCTATGATGTGCTGCGAATGTACACACAGACCATGACGTGGCC 216

Db |||||||

21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40

QY 217 GTGCTCGTGGGAGACCTGAAGCTGCTCAATGAACACCGCTGCTGCTCTGTTGAT 276

Db |||||||

41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60

QY 277 GCCATTCGCGCCCTGATCCACTGAAGCAGCAGGAGGATATGATCAGCTGACCCCGCG 336

Db |||||||

61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80

QY 337 CGCTCCAGGAAGCTGAAGAGGTGCTGTGACCGCTGACCCGAGAGCTCGGCCCTG 396

Db |||||||

81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100

QY 397 AGTGTGCGTGGTGGCTGGAGTTGGCTGTGGGCTTTCATCTCCACCATCATCAAGGC 456

Db |||||||

101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120

QY 457 GGTACGGCAGACAGCGTCCGGCTCCAGGTAGGGAGCAGATCGTCCGGATCAATGGATAT 516

Db |||||||

121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140

QY 517 TCCATCTCCTCTGTACCCATGAGGAGGTCACTCACTTCATTCGACCAAGAAACTGTG 576

Db |||||||

141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160

QY 577 TCCATCAAGTGAACACATCGGCCCTGATCCCGTGAAGACTCTCTGATGAGCCCTC 636

Db |||||||

161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180

QY 637 ACTTGGCAGTATGTGATCAGTTTGTCTCGGAATCTGGGGGCGTGGAGGACCTGGGC 696

Db |||||||

181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200

QY 697 TCCCTCGAATCGGAAACACAGAGAAAGAGTCTTTCATCGACCTGGTAGCTCCCGA 756
|||||
Db 201 SerProGlyAsnArgGluAsnLysGluLysLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCTTGGCTGACGATTTCCACGGCGCCCATCCACAGCCCTGGCATCTTTATCAGCCAT 816
|||||
Db 221 GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTCAAACTGGCTGCCCTGCTGCTGAGTGGGATTGGAGATAGGGACCATGTTGCGAA 876
|||||
Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCAATGGCTGACATCTCTACCTGGATCACAAGAGGCTGTAATGTGCTGAAAAAT 936
|||||
Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
QY 937 AGCGGACGCTGACCATCTCATTGTAGCTGCAGCTGGCGGAGCTGTTTCATGACAGAC 996
|||||
Db 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGAGCGGCTGCGACAGCGCGCGCTGAGCTGCAGCGGAGGAGCTTCTCATGCGAG 1056
|||||
Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuMetGln 320
QY 1057 AAGCGCTGCGATGAGTCCACAGATCTCCAGGACGACGAGATGGAGCGGCAA 1116
|||||
Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln 340
QY 1117 AGGAGAAACAAATGCCCCAGAGGACGACAGAGAAATGAGAGATACCGAAGGAGATG 1176
|||||
Db 341 ArgArgGlyGluIleAlaGlnLysAlaAlaGluGlnArgGlyArgLysGluMet 360
QY 1177 GAACAGATTGTAGAGGAGGAGAAAGTTTAAGAAGCAATGGGAAGAAGCTGGGCTCA 1236
|||||
Db 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluLysTrpGlySer 380
QY 1237 AAGAACAGCTACTTTCCTGCTAAACATCAGCTGAGTACACCCAGTACCCCTCGC 1296
|||||
Db 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCAAGTATGATCAGGAGTGGACCTGAGCTGAGCGCCGACATGACCTGGATGA 1356
|||||
Db 401 LysProLysTrpAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
QY 1357 GGCACGAGGAGGAGGAGGAGGAGGATTTCCGGAATATATAGGAAGGCTTTGACCCCTAC 1416
|||||
Db 421 GlyThrGluGluGlnGlyGluGlnAspPheArgLysTrpGluGluGlyPheAspProTyr 440
QY 1417 TCTATGTCACCCAGACAGATCATGGGAAGGATGTCGGCTCCTACCATCAAGAAG 1476
|||||
Db 441 SerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLys 460
QY 1477 GAGGATCTTACACCTGCGCCCTGGAGGCGGTGGACTCCCCCATGGGAAGGTGGTC 1536
|||||
Db 461 GluGlySerLeuAspLeuAlaLeuGluGlyGlyValAspSerProIleGlyLysValVal 480
QY 1537 GTTTCCTGCTGTATGACGGGGAGCTGCTGACGGGATGTCGTCATGGAAGGGGAC 1596
|||||
Db 481 ValSerAlaValTrpGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAsp 500
QY 1597 GAGATCATGCAATCAACGCGCAAGATTGTGACAGACTACACCTGGCTGAGGCTGACGCT 1656
|||||
Db 501 GluIleMetAlaIleAsnGlyLysIleValThrAspTrpThrLeuAlaGluAlaAspAla 520
QY 1657 GCGCTGAGAGCGCTGGAATCAGGGCGGGGACTGATGACCTGTTGGTGGCTGCTGCTG 1716
|||||
Db 521 AlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValValAlaValCys 540
QY 1717 CCCCCAAGGAGTATGACGATGAGCTGACCTTCTGCTGAAGTCCAAAGGGGAAACCAA 1776
|||||
Db 541 ProProLysGluTrpAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnGln 560

QY 1777 ATTACGCGTTAGGAAACAGTGAGCTCCGGCCCGACCTCGTGAACACAAAGCCTCGGACC 1836
|||||
Db 561 IleHisAlaLeuGlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgThr 580
QY 1837 AGCCTTGAGAGGCGCACATGACACACACAGATGGCATCTTGGGACCTGAATCTATCA 1896
|||||
Db 581 SerLeuGluArgGlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSer 600
QY 1897 CCCAGAACTCAAACTCCCTTTGGCCCTGAACACGAGCCAGATAAAGAACAGCTCGGC 1956
|||||
Db 601 ProArgAsnLeuLysLeuProLeuAlaLeuAsnGlnGlyGlnIleArgAsnSerSerGly 620
QY 1957 CACTTTTGAAGGCAATGTGGAGGAAGGAGGAGGAGGAGCCGCTTTGGGAGAAATCTC 2016
|||||
Db 621 HisPhePheGluGlyGlnCysGlyGlyLysGlyAlaAlaSerArgLeuGlyGluAspLeu 640
QY 2017 AAGGATCCAGACTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2052
|||||
Db 641 LysAspProAspSerHisSerPheProLeuAlaGln 652
RESULT 3
AA07098
ID AA07098 standard; Protein; 481 AA.
XX
AC AA07098;
XX
DT 02-JUL-1999 (first entry)
XX
DE Colon cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN W09904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfeundschoh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Disclosure: Page 664-665; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.

XX
 SQ Sequence 481 AA;

Alignment Scores:

Pred. No.: 2,41e-188 Length: 481
 Score: 2331.00 Matches: 462
 Percent Similarity: 96.06% Conservative: 1
 Best Local Similarity: 95.8% Mismatches: 0
 Query Match: 56.58% Indels: 19
 DB: 20 Gaps: 1

US-09-502-945-4 (1-2236) x AAY07098 (1-481)

QY 97 ATGGACGAAAGTGGCCGAGAAATCCGGCATAAAGTGGATTCTCTGATTGAAATGAT 156
 DB 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuLeuGluAsnAsp 20
 QY 157 GCAGAGAAGACTATCTCTATGATGTGCTCGGAATGTACCCACAGACCACCTGACGTGGCC 216
 DB 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
 QY 217 GTGCTGTGGGAGACTGAAGCTGGTCATCAATGAACACCGCGTGTGCTCTGTTTGTAT 276
 DB 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
 QY 277 GCCATTGGCCGCTGATCCACTGAAGCACACCGAGTGGAAATATGATCAGCTGACCCCGCG 336
 DB 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 QY 337 CGCTCCAGGAAGCTGAAGAGGTGCTGTGGACCGTCTGCACCCGAGGCCCTCGGCCTG 396
 DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
 QY 397 AGTGTGCTGGTGGCTGGAGTTGGCTGTGGCTCTTATCTCCACCTCATCAAGGC 456
 DB 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuLysGly 120
 QY 457 GGTAGCGACAGCGCTCGGGTCCAGGTAGGAGGAGAGTCTCCGGATCAATGATAT 516
 DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 QY 517 TCCATCTCTCTGTACCATGAGGAGTCTCAACTCATTCGGAACCAAGAAACTGTG 576
 DB 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysThrVal 160
 QY 577 TCCATCAAAAGTGAGACACATCGCCGTGATCCCGTGAAAGCTCTCTGATGAGCCCTC 636
 DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
 QY 637 ACTTGGCAGTATGTGATCAGTTGTGTGCGAATCTGGGGGTGCGAGGACGCTGGGC 696
 DB 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200
 QY 697 TCCCTTGGAAATCGGAAACAGGAGAGAGGTCTTCATCAGCCTGTAGCTCCCGA 756
 DB 201 SerProGlyAsnArgGluAsnLysGluLysLysValPheIleSerLeuValGlySerArg 220
 QY 757 GGCTTGGCTGCGAGCATTTCCAGCGGCCCATCCAGACCTGGCATCTTTATCAGCCAT 816
 DB 221 GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
 QY 817 GTGAACCTGGTCCCTGCTGTGAGTGGGATGGAGATAGGGAGCAGAGTTGTCGAA 876
 DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
 QY 877 GTCAATGGCTGACATCTCTTAACCTGGATCAAGAGGCTGTAAATGTCTGTAATAAT 936
 DB 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280

QY 937 AGCCGAGCGCTGACCATCTCCATTTAGTGTGCGAGTGGCCGGAGCTGTTCATGACAGAC 996
 DB 281 SerArgSerLeuThrIleSerIleValAlaAlaAlaGlyArgGluLeuPheMetThrAsp 300
 QY 997 CGGAGCGGCTGGCAGAGCGCGGCGAGCTGAGCTGCAGCGCAGGAGCTTCTCATGAC 1056
 DB 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
 QY 1057 AAGCGCTGGCGATGGAGTCCCAACAGATCTCCAGGAGCAGCAGAGATGGAGCGGCAA 1116
 DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnMetGluArgGln 340
 QY 1117 AGGAGAAAGAAATTTGCCAGAGGAGCAGCAGAGAAATGAGAGATACCGGAGGAGATG 1176
 DB 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
 QY 1177 GAACAGATTGTAGAGGAGGAGAGAGAGTATTAGAGCAATGGGAAGAGACTGGGGCTCA 1236
 DB 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380
 QY 1237 AAGGAACAGCTACTCTTGCCTAAAACCATCTGCTGAGGTACACCCAGTACCCCTTGC 1296
 DB 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 QY 1297 AAGCCAAAGATGATCAGGGAGTGAACCTGAGCTGAGCCCGCAGAGTACCTGGATGGA 1356
 DB 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
 QY 1357 GCGCAGGAGGAGGAGGAGCAGG- - - - - 1381
 DB 421 GlyThrGluGluGluGlyGlu-GlnProGlnGluMetLeuLysArgMetValTyrGln 440
 QY 1382 - - - - - ATTCGGAATATAGGAAGGCTTTGACCCCTACTCTATG 1422
 DB 440 nAspSerIleGlnAspLysIleSerGlyAsnMetArgLysAlaLeuThrProThrLeuCy 460
 QY 1423 TTCACCCAGAGCAGCATCATGGGAAGGATGTCGGGCTCTTACGCATCAAGAAGGAGGA 1482
 DB 460 sSerProGlnSerArgSerTrpGlyArgMetSerGlySerTyrAlaSerArgArgAs 480
 QY 1483 TCTT 1486
 DB 480 pPro 481
 RESULT 4
 AAY07099
 ID AAY07099 standard; Protein; 521 AA.
 XX
 AC AAY07099;
 XX
 DT 02-JUL-1999 (first entry)
 XX
 DE Colon cancer associated antigen precursor sequence.
 XX
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.
 XX
 OS Homo sapiens.
 XX
 PN W09904265-A2.
 XX
 PD 28-JAN-1999.
 XX
 PF 15-JUL-1998; 98WO-US14679.
 XX
 PR 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX

PA (LUDW-) LUDWIG INST CANCER RES.

XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfeundschoh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX WPI: 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX

PS Disclosure: Page 666-667; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

XX Sequence 521 AA;

SQ

Alignment Scores:
Pred. No.: 1.23e-186 Length: 521
Score: 2311.00 Matches: 462
Percent Similarity: 88.70% Conservative: 1
Best Local Similarity: 88.51% Mismatches: 0
Query Match: 56.09% Indels: 59
DB: 20 Gaps: 1

US-09-502-945-4 (1-2236) x AAY07099 (1-521)

QY 97 ATGGACCGAAAGTGCCCGAGAAATCCCGCATAGGTGGATTCTGTGATTAATGAT 156
DB 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
QY 157 GCAGAGAAGGACTATCTCTATGATGCTGCGAATGTACACAGACCATGACGTGGCC 216
DB 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
QY 217 GTGCTCGTGGGAGACTGAAGCTGGTCATCAATGAACCCAGCGCTGCTCTGTTTGTAT 276
DB 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
QY 277 GCATTCGGCCGCTGATCCCACTGAGCAGCAGGTGGATATGATCAGCTGCCCGCG 336
DB 61 AlaileArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
QY 337 CGCTCCAGGAAGCTGAAGGAGTGCCTCTGACCGCTGCACCCGAGGCGCTCGGCCCTG 396
DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
QY 397 AGTGTGGTGGTGGCTGGAGTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGGC 456
DB 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuLysGly 120
QY 457 GGTCAAGCAGACAGCTGGGCTCCAGGTAGGGGAGCAGATCGTCCGGATCATGATAT 516
DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
QY 517 TCCATCTCCCTCTGATCCCATGAGGAGTTCATCACTCATTCGAAACCAAGAACTGTG 576
DB 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleLeuThrLysLysThrVal 160

QY 577 TCATCAAAAGTGAGACACATCGCCCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTC 636
DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
QY 637 ACTTGGCAGTATGTGGATCAGTTGTCTCGAAATCTGGGGCGGTGGAGCAGCCTGGGC 696
DB 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200
QY 697 TCCCTCGAAATCGGAAAAACAAGGAGAAGAGTCTTCATCAGCCCTGTAGGCTCCCGA 756
DB 201 SerProGlyAsnArgGluAsnLysGluLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCCTTGGCTGCAGCATTTCCAGGGGCCCATCCAGAGCCTGGCATCTTTATCAGCCAT 816
DB 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTGAAACCTGGCTCCCTCTGTCTGAGTGGGATTGAGATAGGGACACAGATTGTGCAA 876
DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCAATGGCTGCGACTTCTTAACCTGGATCACAAGAGGCTGTAAATGTGCTGAAAAAT 936
DB 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
QY 937 AGCCGACGCTGACCATCTCCATTGTAGTCCAGCTGGCCGGGAGCTGTTTCATGACAGAC 996
DB 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGAGGCGCTGGCAGAGGCGCGCAGCTGAGCTGAGCGGCGAGGAGCTTCTCATGCAG 1056
DB 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
QY 1057 AAGCGCTGGCGATGGATGCCAACAGATCTCCAGAGCAGCAGGAGATGGAGCGCAA 1116
DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln 340
QY 1117 AGGAGAAAGAAATGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1176
DB 341 ArgArgLysGluIleAlaGlnLysAlaGlnLysAlaGlnLysAlaGlnLysAlaGln 360
QY 1177 GAACAGATTGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1236
DB 361 GluGlnIleValGluGluGluGluLysPheLysLysGlnTrpGluLysTrpGlySer 380
QY 1237 AAGCAACAGCTACTCTTTCCTTAAACCATCATCTCTGAGGTACACCCAGTACCCCTTGC 1296
DB 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCAAAAGTATGATCAGGAGTGAACCTGAGCTCGAGCCCGCAGATGACCTGATGGA 1356
DB 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
QY 1357 GGCACGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1381
DB 421 GlyThrGluGluGlnGlyGlu-GlnThrPheCysProSerProGlnProProArgGlyPr 440
QY 1381 ----- 1381
DB 440 GglyValSerThrIleSerLysProValMetValHisGlnGluProAsnPheIleTyrAr 460
QY 1381 ----- 1381
DB 460 gProAlaValLysSerGluValLeuProGlnGluMetLeuLysArgMetValValTyrGl 480
QY 1382 -----ATTTCGGAAATATAGGAAGGCTTTGACCCCTACTCTATG 1422
DB 480 nAspSerIleGlnAspLysIleSerGlyAsnMetArgLysAlaLeuThrProThrLeuCy 500
QY 1423 TTCACCCAGCAGACATCATGGGAAGGATGTCCGGCTCTACCGCATCAAGAGAGGGA 1482
DB 500 sSerProGlnSerArgSerTrpGlyArgMetSerGlySerTyrAlaSerArgArgAs 520
QY 1483 TCCT 1486

```

Db 520 pPro 521
|||||
RESULT 5
RAY07095
ID AAY07095 standard; Protein; 403 AA.
XX AC AAY07095;
XX DT 02-JUL-1999 (first entry)
XX DE Colon cancer associated antigen precursor sequence.
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX OS Homo sapiens.
XX PN WO9904265-A2.
XX PD 28-JAN-1999.
XX PF 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX (LUDW-) LUDWIG INST CANCER RES.
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
XX Pfeundschoh M, Sahin U, Scanlan MJ, Stockert E;
XX Tureci O;
XX WPI; 1999-132448/11.
XX New isolated cancer associated nucleic acids and polypeptides -
XX isolated using sera from cancer patients, used to develop products
XX for the diagnosis, monitoring or treatment of cancers
XX Disclosure; Page 659-660; 787pp; English.
XX The invention relates to a method for diagnosing a disorder characterised
XX by expression of a human cancer associated antigen precursor coded for by
XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX biological sample isolated from a subject with an agent that specifically
XX binds to the NAM, an expression product or a fragment of an expression
XX product complexed with an HLA molecule; and (b) determining the
XX interaction between the agent and the NAM or the expression product as a
XX determination of the disorder. The products and methods can be used in
XX the diagnosis, monitoring, research, or treatment of conditions
XX characterised by the expression of various cancer associated antigens.
XX The invention provides nucleic acid sequences and encoded polypeptides
XX which are cancer associated antigen precursors expressed in human breast
XX cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX lung cancer.
SQ Sequence 403 AA;
Alignment Scores:
Pred. No.: 2,9e-164 Length: 403
Score: 2046.00 Matches: 403
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 49,66% Indels: 0
DB: 20 Gaps: 0
US-09-502-945-4 (1-2236) x AAY07095 (1-403)

```

```

QY 97 ATGAGCCGAAAGTGGCCCGGAGAAATTCGGGCATAGGTGGATTTCTGATTGAAATGAT 156
Db 1 MetAspArgLysValAlaAargGluPheArgHisLysValAspPheLeuLeuGluAsnAsp 20
QY 157 GCAGAGAGGAGTATCTCTATGATGCTGCGAATGTACCAACAGACCATGAGTGGCC 216
Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetCaspValAla 40
QY 217 GTGCTGCTGGGAGACCTGTAAGCTGGTCAATCAATGAACCCAGCCGCTGCTCTTTGAT 276
Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
QY 277 GCCATTGGCCGCTGATCCCACTCAAGCACCAGGTGAATATGATGATGACCTGACCCCGG 336
Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
QY 337 CGCTCCAGGAAGCTGAAGGAGGTGCTGTGGACCGCTGACACCCGAGGCTCGGCGCTG 396
Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
QY 397 AGTGTGCTGCTGGCTGGAGTTTGGCTGGGCTCTTCATCTCCCACTCATCAAAAGGC 456
Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
QY 457 GGTCAAGCAGACAGCTCGGCTCCAGGTAGGGAGCAGATCGTCCGGATCAATGGATAT 516
Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
QY 517 TCCATCTCCTCCTGTACCCATGAGGAGGTCTCAACCTCATTCGAACCAAGAAACCTGTG 576
Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
QY 577 TCCATCAAGTAGACACATCGGCTGATCCCGCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTC 636
Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
QY 637 ACTTGGCAGTATGTGGATCAGTTTGTGCGAATCTGCGGCGCTGGGAGCAGCCTGGGC 696
Db 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlySerLeuGly 200
QY 697 TCCCTCGAATCGGGAACAAAGAGGAGAAAGTCTTTCATCAGCCTGTGTAGCTCCCGA 756
Db 201 SerProGlyAsnArgGluAsnLysGluLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCTTGGCTGACGATTTCCAGCGGCCCATCCAGAGCCTGGCATCTTTATCAGCCAT 816
Db 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTAAACCTGGCTCCCTGCTGCTGAGTGGGATGGAGATAGGGACAGATTTGCGAA 876
Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCAATGGCTGACCTTCTTAACCTGGATCAAGAGGCTGTAATGTGTGTAAGAAAT 936
Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
QY 937 AGCCGAGCCTGACCATCTCCATTGTAGCTGACGCTGGCCGGAGCTGTTCATGACAGAC 996
Db 281 SerArgSerLeuThrIleSerIleValAlaAlaAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGAGCGGCTGGCAGCGGCGGAGCTGTAGCTGCAGCGGAGGAGCTTCTTCATGCGAG 1056
Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuLeuMetGln 320
QY 1057 AAGCGGCTGGGATGGAGTCCACACAGATCCCTCCAGGACGACGAGAGATGGAGCGGCAA 1116
Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln 340
QY 1117 AGGAGAAAGAAATTTGCCAGAGGAGGACGAGAGAAATAGAGATATCCGGAGAGATG 1176
Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
QY 1177 GAACAGATTGTAGAGGAGGAGAGAAAGTTTAAAGAACCAATGGGAAGAGACTGGGCTCA 1236

```



```

Db 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380
QY 1237 AAGGAACAGCTACTTGGCTTAAACCATCTGCTGAGGTACACCCAGTACCCCTTGC 1296
Db 381 LysGluGlnLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCCAAAG 1305
Db 401 LysProLys 403
RESULT 6
AA07097
ID AAY07097 standard; Protein; 328 AA.
XX
AC AAY07097;
XX
DT 02-JUL-1999 (first entry)
XX
DE Colon cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN W09904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Disclosure; Page 662-663; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 328 AA;
Alignment Scores:
Pred. No.: 9.79e-138 Length: 328

```

```

Score: 1732.00 Matches: 328
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.04% Indels: 0
DB: 20 Gaps: 0
US-09-502-945-4 (1-2236) x AAY07097 (1-328)
QY 1069 ATGGAGTCCCAACAGATCTCCAGAGCAGCAGAGTGGAGCGCAAGAGAGAAAGAA 1128
Db 1 MetGluSerAsnLysIleLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 20
QY 1129 ATTGCCAGAGGAGCAGAGAGAAATGAGAGATACCGGAAGAGGATGGAACAGATTCTA 1188
Db 21 IleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMetGluGlnIleVal 40
QY 1189 GAGGAGGAAGAGAGTTTAAGAACATGGGAAGAAAGACTGGGGCTCAAGAGACAGCTA 1248
Db 41 GluGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySerLysGluGlnLeu 60
QY 1249 CTCTTGGCTAAACCATCTGAGTACACCCAGTACCCCTTGGCAAGCCAAAGTAT 1308
Db 61 LeuLeuProLysThrIleThrAlaGluValHisProValProLeuArgLysProLysTyr 80
QY 1309 GATCAGGAGGTGGAACTGAGCTCGAGCCCGCAGATGACCTGGATGGAGGACGAGGAG 1368
Db 81 AspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGlyGlyThrGluGlu 100
QY 1369 CAGGAGAGCAGGATTTCGGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTTCCAC 1428
Db 101 GlnGlyGluGlnAspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThr 120
QY 1429 CCAGAGCAGATCATGGGAAGGATGTCGGCTCTTACCATCAAGAGAGGAGGATCTTTA 1488
Db 121 ProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeu 140
QY 1489 GACTTGGCCCTGGAAAGCGGTGTGAGCTCCCATTTGGGAAGGTGGTGGTTCTGCTGTG 1548
Db 141 AspLeuAlaLeuGluGlyGlyValAspSerProIleGlyLysValValValSerAlaVal 160
QY 1549 TATGAGCGGGGAGCTGCTGAGCGGCATGTGGCATGTGGAAGGGGAGGAGATCATGGCA 1608
Db 161 TyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAla 180
QY 1609 ATCAACGGCAAGATTGTGACAGACTACACCTCGGTGAGGCTGACGCTGCGCTGCAGAG 1668
Db 181 IleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAlaLeuGlnLys 200
QY 1669 GCCTGGAATCAGGCGGGGAGTGGATCGACCTTGTGGTGGTGGTGGTGGTGGTGGTGG 1728
Db 201 AlaTrpAsnGlnGlyAspTrpIleAspLeuValValAlaValAlaValCysProLysGlu 220
QY 1729 TATGACGATGAGCTGACCTTCTTGTGAAGTCCAAAGGGGAAACCAATTCACCGCTTA 1788
Db 221 TyrAspAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnGlnIleHisAlaLeu 240
QY 1789 GGAACAGTGAAGTCTCGGCGCCACCTCGTGACACAAAGCCCTCGGACAGCCTTGAGAGA 1848
Db 241 GlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgThrSerLeuGluArg 260
QY 1849 GGCCACATGACACACACAGATGGATCCTTGGGACCTGAATCTATCACCAGGATCTC 1908
Db 261 GlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSerProArgAsnLeu 280
QY 1909 AAATCCCTTGGCCCTCAACCGCCAGATAGGAACAGCTCGGCGCACCTTTTGTAA 1968
Db 281 LysLeuProLeuAlaLeuAsnGlnGlnIleArgAsnSerSerGlyHisPheGlu 300
QY 1969 GGCCAAATGTGGAGAAAGGAGCAGCCAGCCGTTGGGAGAGATCTCAAGATCCAGAC 2028
Db 301 GlyGlnCysGlyGlyGlyAlaAlaSerArgLeuGlyGluAspLeuLysAspProAsp 320
QY 2029 TCTCATTCCTTTCTCTGGCCAG 2052

```

```

Db 321 SerHisSerPheProLeuAlaGln 328
RESULT 7
AAY38440
ID AAY38440 standard; Protein; 268 AA.
XX
AC AAY38440;
XX
XX 18-OCT-1999 (first entry)
XX
DE Human secreted protein encoded by gene No. 5.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
XX WO9935158-A1.
XX
XX 15-JUL-1999.
XX
XX 06-JAN-1999; 99WO-US00108.
XX
XX 07-JAN-1998; 98US-0070704.
XX
XX 07-JAN-1998; 98US-0070657.
XX
XX 07-JAN-1998; 98US-0070658.
XX
XX 07-JAN-1998; 98US-0070692.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;
XX Olsen HS, Rosen CA, Ruben SM, Soppet DR;
XX
XX WPI; 1999-444190/37.
XX
XX N-PSDB; AA206223.
XX
XX New isolated human genes and the secreted polypeptides they encode
XX
XX Disclosure; Page 206-207; 227pp; English.
XX
XX This sequence represents a secreted human protein encoded by the gene
XX clone detailed in the descriptor line. The gene can be used to generate
XX fusion proteins by linking to the gene to a human immunoglobulin FC
XX portion (e.g. AA206210) for increasing the stability of the fused
XX protein as compared to the human protein only.
XX The invention relates to 36 novel genes and their fragments (nucleic
XX acid sequences: AA206219-206263; amino acid sequences AAY38386-Y38498)
XX which are useful for preventing, treating or ameliorating medical
XX conditions e.g. by protein or gene therapy. Also, pathological
XX conditions can be diagnosed by determining the amount of the new
XX polypeptides in a sample or by determining the presence of mutations in
XX the new polynucleotides. Specific uses are described for each of the 36
XX polynucleotides, based on which tissues they are most highly expressed in
XX (see AA206219 for described uses).
XX
XX Sequence 268 AA;
XX
Alignment Scores:
Pred. No.: 1.88e-105 Length: 268
Score: 1350.00 Matches: 262
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 32.77% Indels: 0
DB: 20 Gaps: 0
XX
US-09-502-945-4 (1-2236) x AAY38440 (1-268)

```

```

QY 964 GCTGCAGCTGGCGGAGCTGTTTCATGACAGACCGGGAGCGGTGGCAGAGCGCGGAG 1023
Db 6 AlaAlaAlaGlyArgGluLeuPheMetThrAspArgGluLeuAlaGluAlaArgGln 25
QY 1024 CGTGAGCTGCAGCGGAGGAGCTTTCATGACAGAACGGCTGGCGAGTCCACAAAG 1083
Db 26 ArgGluLeuGlnArgGlnGluLeuLeuMetGlnLysArgGluLeuAlaMetGluSerAsnLys 45
QY 1084 ATCTCTCCAGGAGCAGGAGGATGGAGCGGCAAGAGGAGAAAGAAATTTGCCAGAGGCA 1143
Db 46 IleLeuGlnGlnGlnGlnGlnMetGluArgGlnArgArgLysGluLeuAlaGlnLysAla 65
QY 1144 GCAGAGAAATAGAGATATCCGGAAGGAGATGGACAGATTTAGAGGAGGAGAAG 1203
Db 66 AlaGluGluAsnGluArgTyrArgLysGluMetGluGlnIleValGluGluGluLys 85
QY 1204 TTTAAGAGCAATGGGAAGAGACTGGGGCTCAAGAGACAGCTACTCTTGCCTAAACC 1263
Db 86 PheLysLysGlnTrpGluGluAspTrpGlySerLysGluGlnLeuLeuLeuProLysThr 105
QY 1264 ATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATGATCAGGAGTGGAA 1323
Db 106 IleThrAlaGluValHisProValProLeuArgLysProLysTyrAspGlnGlyValGlu 125
QY 1324 CTTGAGCTGAGCCCGCAGATGACCTGGATGGAGGACGAGGAGGAGGAGGAGGAT 1383
Db 126 ProGluLeuGluProAlaAspAspLeuAspGlyGlyThrGluGluGlnGlyGluGlnAsp 145
QY 1384 TTCCGGAAATATCAGGAAGGCTTTGACCCCTACTCTATGTTTACCCCGCAGAGCATCG 1443
Db 146 PheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGluGlnIleMet 165
QY 1444 GGAAGGATGTCGGCTCTACGCATCAAGAGAGGAGGATCCTTAGACCTGGCCCTGGAA 1503
Db 166 GlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeuAspLeuAlaLeuGlu 185
QY 1504 GCGGTGTGACTCCCGCATTGGGAAGGTGGTGGTTCCTGTGTGTATGAGGGGAGCT 1563
Db 186 GlyGlyValAspSerProIleGlyLysValValValSerAlaValTyrGluArgGlyAla 205
QY 1564 GCTGAGCGCATGCTGGCATTTGGAAGGGGAGGAGATCATGCCATCAACGCCAGATT 1623
Db 206 AlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLysIle 225
QY 1624 GTGACAGACTACACCTGCTGAGCTGAGCTGCGCTGCAGAGGCGCTGGAATCAGGCG 1683
Db 226 ValThrAspTyrThrLeuAlaGluAlaAspAlaAlaLeuGlnLysAlaTrpAsnGlnGly 245
QY 1684 GGGGACTGGATCAGCTTGTGGTTCGCTGCGCCCGCAGAGGAGTATGACGATGCTG 1743
Db 246 GlyAspTrpIleAspLeuValValAlaValCysProLysGluTyrAspAspGluLeu 265
QY 1744 ACCTTC 1749
Db 266 ThrPhe 267
XX
XX RESULT 8
XX AAB53238
XX ID AAB53238 standard; Protein; 120 AA.
XX
XX AC AAB53238;
XX
XX DT 09-MAR-2001 (first entry)
XX
XX Human colon cancer antigen protein sequence SEQ ID NO:778.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX identification; cytostatic; cardioactive; neuroprotective; vulnary;
XX immunomodulatory; muscular; gynaecological; gastrointestinal;
XX nephrotropic; antiinfective; antibacterial; gene therapy; wound;
XX neural disorder; immune system disorder; muscular disorder;
XX reproductive disorder; gastrointestinal disorder; renal disorder;

```

KW infectious disease; cardiovascular disorder.
XX Homo sapiens.
XX WO200055351-A1.
XX PD 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US05883.
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX WPI; 2000-587534/55.
XX DR N-PSDB; AAC97995.
XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
XX Claim 11; Page 1339; 2104pp; English.
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
XX invention.
XX
XX SQ Sequence 120 AA;
Alignment Scores:
Pred. No.: 3.33e-29 Length: 120
Score: 447.00 Matches: 92
Percent Similarity: 87.04% Conservative: 2
Best Local Similarity: 85.19% Mismatches: 4
Query Match: 10.85% Indels: 10
DB: 21 Gaps: 2
US-09-502-945-4 (1-2236) x AAB53238 (1-120)
QY 1315 GGAGTGAACCTGAGCTCGAGCGCGCAGATGACCTGGATGAGGACGAGGAGGAGGGA 1374
Db 23 GlyLeuArgProValLeuGlnPro-----ArgGlnGly 33
QY 1375 GAGCAGGATTCGGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTTACCCCGAG 1434
Db 34 ---GlnAspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGlu 52
QY 1435 CAGATCATGGGGAAGGATGTCCGCTCTACGCATCAAGAAGGAGGATCCTTAGACCTG 1494
Db 53 GlnIleMetGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeuAspLeu 72
QY 1495 GCCCTGAAGCGGTGTGGACTCCCATTTGGGAAGTGTCTTTCTGCTGTGTATGAG 1554
Db 73 AlaLeuGluGlyGlyValAspSer**IleGlyLysValValValSerAlaValTyrGlu 92
QY 1555 CGGGGACCTGCTGAGCGGCATGTGGCATGTGAAAGGGACGAGATCATGGCAATCAAC 1614
Db 93 ArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsn 112

QY 1615 GGCAAGATTGTGACAGACTACACC 1638
Db 113 GlyLysIleValThrAspTyrThr 120
RESULT 9
ABB61924
ID ABB61924 standard; Protein; 493 AA.
XX AC ABB61924;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 12564.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL06027.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 12564; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 493 AA;
Alignment Scores:
Pred. No.: 4.24e-27 Length: 493
Score: 425.50 Matches: 128
Percent Similarity: 42.76% Conservative: 58
Best Local Similarity: 29.43% Mismatches: 164
Query Match: 10.33% Indels: 85
DB: 22 Gaps: 11
US-09-502-945-4 (1-2236) x ABB61924 (1-493)
QY 340 TCCAGAGACTGAAGAGGTGGTGTGACCGCTGTGCACCCCGAAGGCGTC----- 390
Db 58 SerSerArgLeuArgValLeuArgLeuValArgProHisGlnArgArgLeuValGly 77
QY 391 -----GGCCTGAGTGTGCGTGGTGGCTGAGTTGCGCTGT 426
Db 78 GlyProGluArgGlySerThrTyrGlyPheThrValArgGlyArgGluHisGlyThr 97

QY	427	GGGCTCTTCACTCCACCTCATCAAAGCGGTGCAGGCACACAGCCTCGGCTCCAGTA	480
Db	98	GlyPheValSerHisValGluHisGlyGlyGluAlaHisLeuLysGlyLeuArgIle	117
QY	487	GGGACGAGATCGTCGGATCAATGGATATTCCATCTCCTTGTTACCCTCAGGAGTCT	546
Db	118	GlyAspGlnIleLeuArgIleAsnGlyPheArgLeuAspAlaValHisLysGluPhe	137
QY	547	ATCAACCTCATCGAACCAAACTGTGTCTCCATCAAAGTGAGACACATCGGCTGATC	606
Db	138	IleGlnLeuValAlaGlyGlnAspArgValThrLeuLysValArgGlyValGlyMetLeu	157
QY	607	CCCTGAAAAGCTCTCTGATGAGCCCTCACATGGCAGTATGTGGATCAGTTTGTGTCG	666
Db	158	ProValArgAspLeuProGluGluArgLeuSerTrpSerValLys-----LeuPro	175
QY	667	GAATCTGGGGGTGGAGCGACCTGGGCTCCCCTGGAAATCGGAAACAAGAGAGAAG	726
Db	176	SerValSerGlyThrProSerGluSerSerPheLysGlyGluArgArgGlyAlaSerArg	195
QY	727	AAGTCTTTCATCAGCTG---GTAGCTCCCGA---GGCTTGGCTGCAGCATTTCCACG	780
Db	196	AspIleSerValLeuLeuHisValAlaProArgThrLysLeuGlyLeuGlyIleCysLys	215
QY	781	GGCCCCATCCAGAAGCTCGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGCTGCT	840
Db	216	GlyProGluTrpLysProGlyIlePheValGlnPheThrLysAspArgSerValAlaArg	235
QY	841	GAGTGGGATGGAGATAGGGGACCAAGATGTGGAAGTCAATGGCTGACTTCTCTAAC	900
Db	236	GluAlaGlyLeuArgProGlyAspGlnIleLeuSerValAsnSerIleaspPheSerAsp	255
QY	901	CTGGATCACAGGAGCTGTAATGTGCTGAAAAATAGCCGAGCTGCACATCTCCATT	960
Db	256	ValLeuPheSerGluAlaValAlaValMetLysSerSerLysLeuAspMetValVal	275
QY	961	GTAGCTGACCTGCCGGGAGCTGTTTCATGACAGACGGGACGGCTGCAGAGCGCGG	1020
Db	276	ArgThrAlaAlaGlyCysAspLeuPheProGlyGlu-SerSerGlyTyrrAsnSerAl	295
QY	1021	CAGGTCAGCTGCAGCGGAGGAGCTTCTCATGCAAGACGGCTGGCATGGAGTCCAAC	1080
Db	295	aSer---SerValThrGlyAspGlnSerProCys-----TpIlaAspAl	309
QY	1081	AAGATCTCCAGCAGCAGGAGGATGGAGCGCAAGAGGAGAAAGAAATTCGCCAGAG	1140
Db	309	aLysSerLysArg-LeuThr-----AlaValargGluGluSerGlyAlaGlyGlyG	326
QY	1141	GCAGCAGAGGAAATGAGAGATACCGGAGGAGATGGACAGATTGTAGAGGAGGAAG	1200
Db	326	lyGlyCysGlyLeuSerSerAlaProGlyAlaGlySerProAsnTrpSerGlnGlyValG	346
QY	1201	AAGTTTAAGAACCAATGGGAAGACTG-----1229	
Db	346	luValHisLysGlnMetAsnLysThrIleIleLysLeuThrgluAsnGlyThrSerIleA	366
QY	1230	-----GGGCTCAAGGAA	1242
Db	366	snAsnThrTyrlleAlaSerThrGlyGlySerSerValSerGlySerGlySerThrGlyS	386
QY	1243	CAGCTACTCTTGCTATAAACCATCATCTGCTAGGTACACCCAGTACCCCTTCGCAAGCA	1302
Db	386	erGlyThrSerGlyArgSerGlnInSerGlnSerAsnProSerAsnProSerArg----	404
QY	1303	AAGTATCATCGGAGTGAACCTGAGCTCGAGCCCGCAGATGACCT-----1349	
Db	405	-----AsnSerThrThrMetLysArgSerHisLeuArgproValAsnSerAlag	421
QY	1350	-----GGATGGAGCAGGAGGAGGAGGAGCAGGATTCCTCGGAATATGAG	1398
Db	421	lySerGlylleGlyLeuSerSerGlySerAlaglySerAlacly-----435	

QY	1399	GAAGCGTTTGACCCCTACTCTATGTTTCACCCCAAGACAGATCATGGGGAGGATGTCGGG	1458
Db	436	-----SerAlaGlySerSerG	441
QY	1459	CTCCTACGCATCAAGAGGAGGAGGATCCTTAGACCTGGCCCT	1499
Db	441	lySerGlySerArgSerGlyGlyValIleAlaProAlaPro	454
RESULT 10			
ID	AAAY38390	standard; Protein; 98 AA.	
AC	AAAY38390;		
DT	30-SEP-1999	(first entry)	
DE	Human secreted protein encoded by gene No. 5.		
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;		
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;		
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;		
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;		
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;		
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;		
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;		
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.		
OS	Homo sapiens.		
PN	WO9935158-A1.		
PD	15-JUL-1999.		
PF	06-JAN-1999; 99WO-US00108.		
PR	07-JAN-1998; 98US-0070704.		
PR	07-JAN-1998; 98US-0070657.		
PR	07-JAN-1998; 98US-0070658.		
PR	07-JAN-1998; 98US-0070692.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;		
PI	Olsen HS, Rosen CA, Ruben SM, Soppet DR;		
DR	WPI; 1999-444190/37.		
DR	N-PSDB; AAZ06223.		
PT	New isolated human genes and the secreted polypeptides they encode		
PS	Claim 11; Page 178-179; 227pp; English.		
CC	This sequence represents a secreted human protein encoded by the gene		
CC	clone detailed in the descriptor line. The gene can be used to generate		
CC	fusion proteins by linking to the gene to a human immunoglobulin Fc		
CC	portion (e.g. AAZ06210) for increasing the stability of the fused		
CC	protein as compared to the human protein only.		
CC	The invention relates to 36 novel genes and their fragments (nucleic		
CC	acid sequences: AAZ06219-206263; amino acid sequences AAAY38386-Y38498)		
CC	which are useful for preventing, treating or ameliorating medical		
CC	conditions e.g. by protein or gene therapy. Also, pathological		
CC	conditions can be sampled or by determining the amount of the new		
CC	polypeptides in a sample or by determining the presence of mutations in		
CC	the new polynucleotides. Specific uses are described for each of the 36		
CC	polynucleotides, based on which tissues they are most highly expressed in		
CC	(see AAZ06219 for described uses).		
SQ	Sequence	98 AA;	
Alignment Scores:			
Pred. No.:	1.05e-26	Length:	98
Score:	417.00	Matches:	88
Percent Similarity:	83.81%	Conservative:	0

Alignment Scores:		
Pred. No.:	1.05e-26	98
Score:	417.00	88
Percent Similarity:	83.81%	0
Length:		
Matches:		
Conservative:		

Best Local Similarity: 83.81% Mismatches: 0
Query Match: 10.12% Indels: 17
DB: 20 Gaps: 1

US-09-502-945-4 (1-2236) x AAY38390 (1-98)

QY 676 GCGTCGAGGAGGCTGGCTCCCTGGAAATCGGGAAACAGAGGAGAAAGTCTTC 735
DB 10 GlyValArgGlySerLeuGlySerProGlyAsnArgGluAsnLysGluLysValPhe 29

QY 736 ATCAACCTGGTGGTCCCGAGGCTCGCTGCAGCATTCCAGCGCCCATCCAGAAG 795
DB 30 IleSerLeuValGlySerArgGlyLeuGlyCysSerIleSerGlyProIleGlnLys 49

QY 796 CTGGCATCTTTATCAGGCATGTGAACCTGGCTCCCTGCTCTGCTGAGGTGGATTGGAG 855
DB 50 ProGlyIlePheIleSerHisValLysProGlySerLeuSerAlaGluValGlyLeuGlu 69

QY 856 ATAGGGACCATGTCGAATCTAATGGCTGCGACTTCTTAACCTGGATCACAAGAG 915
DB 70 IleGlyAspGlnIleValGluValAsnGlyValAspPheSerAsnLeuAspHisLysGlu 89

QY 916 GCTGTAATGCTGTAATAATAGCCGACCTGACCATCTCCATTGTAGCTGCGCTGCG 975
DB 90 -Leu-----GlnLeuAl 93

QY 976 CGGAGCTGTTC 988
DB 93 aGlySerCysSer 97

RESULT 11
AAY38437
ID AAY38437 standard; Protein; 95 AA.
AC AAY38437;
XX
DT 18-OCT-1999 (first entry)
XX
DE Human secreted protein encoded by gene No. 5.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO935158-A1.
XX
PD 15-JUL-1999.
XX
PF 06-JAN-1999; 99WO-US00108.
XX
PR 07-JAN-1998; 98US-0070704.
PR 07-JAN-1998; 98US-0070657.
PR 07-JAN-1998; 98US-0070658.
PR 07-JAN-1998; 98US-0070692.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
XX
DR WPI; 1999-444190/37.
DR N-PSDB; AA206223.
XX
PT New isolated human genes and the secreted polypeptides they encode
XX Disclosure; Page 205; 227pp; English.

XX This sequence represents a secreted human protein encoded by the gene
CC clone detailed in the descriptor line. The gene can be used to generate
CC fusion proteins by linking to the gene to a human immunoglobulin Fc
CC portion (e.g. AAZ06210) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 36 novel genes and their fragments (nucleic
CC acid sequences: AA206219-206263; amino acid sequences: AAY38386-Y38498)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 36
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAZ06219 for described uses).

XX
SQ Sequence 95 AA;

Alignment Scores: 8.81e-26 Length: 95
Pred. No.: 406.00 Matches: 90
Score: 60.13% Conservative: 2
Percent Similarity: 58.82% Mismatches: 1
Best Local Similarity: 9.85% Indels: 60
Query Match: 20 Gaps: 1
DB: 1

US-09-502-945-4 (1-2236) x AAY38437 (1-95)

QY 1752 GCTGAAGTCCAAAGGGGAAACCAATTCACGCGTTAGGAAACAGTCCGCGCCCA 1811
DB 3 SerGluValGlnLysGlyLysProAsnSer-ProLeuGlyAsnSerGluLeuArgProH1 22

QY 1812 CCTCGTGAACACAAAGCCTCGGACGAGCCTTGAGAGAGGCCACATGACACACAGATG 1871
DB 22 sLeuValAsnThrLysProArgThrSerLeuGluArgGlyH1s ----- 36

QY 1872 GCATCCTTGGGACCTGAATCTATCACCAGGAAATCTCAAACTCCCTTTGGCCTGAACCA 1931
DB 36 ----- 36

QY 1932 GGGCCAGATAAGGAACAGCTCGGGCCACTTTTGAAGGCCAATGTGAGGAAAGGAGC 1991
DB 36 ----- 36

QY 1992 AGCCAGCCGTTTGGGAGAAGATCTCAAGATCCAGACTCTCATCTTCTCTGCGCCA 2051
DB 37 -----Thr-IleProPheLeuTrpProS 44

QY 2052 GTGAATTTGGTCTCTCCAGCTTTGGGGGACTTCTTCCTTGAACCTTAATAGACCCAC 2111
DB 44 erGluPheGlyLeuSerGlnLeuTrpGlyThrProSerLeuAsnProAsnLysThrProL 64

QY 2112 TGGAGTCTCTCTCTCCATCCCTCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2171
DB 64 euGluSerLeuSerLeuHisProSerProLeuProSerAlaLeuIleAlaAlaArgIleV 84

QY 2172 TCACCTCCAAACCTTACTCTGAGCTCATTAATAAAA 2206
DB 84 alThrProAsnLeuThrLeuSerSerLeuIleLys 95

RESULT 12
ABG22366
ID ABG22366 standard; Protein; 1037 AA.
XX
AC ABG22366;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22357.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX

OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
XX	
PD	11-OCT-2001.
XX	
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSBQ INC.
XX	
PI	Dzmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	N-PSDB; AAS86553.
XX	
PT	New Isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity
XX	
XX	
PS	Claim 20; SEQ ID No 52725; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful in medical
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences .
XX	
XX	Sequence 1037 AA;
XX	

Percent Similarity:	34.40%
Best Local Similarity:	21.71%
Query Match:	6.59%
DB:	23

Conservative:	83
Mismatches:	194
Indels:	235
Gaps:	28

US-09-502-945-4 (1-2236) x AAE21718 (1-928)

QY	385	GGCTCGGCTGAGTGTGCCTGGCTGAGTTT---GGTGTCGGCTCTTCATCTCC	441
Db	30	GlyPheGlyLeuAlaIleSerGlyArgAspArgProGlyCysMetValValSer	49
QY	442	CACCTCATCAAGGCGGTGACGACAGCGTCCGGCTCCAGTAGGGACGAGATCGTC	501
Db	50	AspValValProGlyGlyProAlaGluGly---ArgLeuGlnThrGlyAspHisIleVal	68
QY	502	CGGATCAATGGAATTATTCATCTCCTCTCTGACCATTACGAGAGTCAACAACCTCATTCGA	561
Db	69	MetValAsnGlyValSerMetGluAsnAlaThrSerAlaPheAlaIleGlnIleLeuLys	88
QY	562	ACC---AAGAAAACTGTGCCATCAAAGTC-----AGCACATCCGGCTGATCCCC	609
Db	89	ThrCysThrLysMetalAsnIleThrValLysArgProArgIleHisLeuProAla	108
QY	610	GTGAAAGACTCTCCT-----CATGAGCCCTC	636
Db	109	ThrLysAlaSerProSerSerProGlyArgGlnAspSerAspCluaAspSpGlyProGln	128
QY	637	ACTTGGCAGTATGTGGATCAG-----TTTGTGTGGAATCTGGG	675
Db	129	ArgValGluGluValAspGlnGlyArgGlyTyraSpGlyAspSerSerGlySerGly	148
QY	676	-----GGCGTCGGAGCGCCGTGGCGCTC	699
Db	149	ArgSerTirpAspGluArgSerArgProArgProGlyArgArgGlyArgAlaGlySer	168
QY	700	CCTGGAATCGGNAACAGGAGAAGAGTCTTCATCAGCTGGTAGGCTCCCGAGGC	759
Db	169	HisGlyArgArgSerProGlyGly-----GlySerGluAla	180
QY	760	CTTGGCTGCAGCATTTCCAGCGC-----CCCATCCAG-----AAG	795
Db	181	AsnGlyLeuAlaLeuValSerGlyPheLysArgLeuProArgGlnAspValGlnMetLys	200
QY	796	CCT-----	798
Db	201	ProValLysSerValLeuValLysArgAspSerGluGluPheGlyValLysLeuGly	220
QY	799	---GGCATCTTTATCAGCCATGTGAACCTGGCTCCTCTCTGCT---CAGGTGGGATTG	852
Db	221	SerGlnIlePheIleLysHisIleThrAspSerGlyLeuAlaAlaArgHisArgGlyLeu	240
QY	853	GAGATAGGGACAGATTGTCGAAGTCANTGGCTCGACTTCTTAACCTGGATCACAAAG	912
Db	241	GlnGluGlyAspLeuIleLeuGlnIleAsnGlyValSerSerGlnAsnLeuSerLeuAsn	260
QY	913	GAGGCTATAATCTGTGAAAATAFSCCGCAGCCTGACCATCTCCATCTAGTGCAGCT	972
Db	261	AspThrArgLeuIleGluLysSerGluGly---LysLeuSerLeuLeuValLeuArg	279
QY	973	GGCCGGGAGCTGTTTCATGCACAG-----	996
Db	280	AspArgGlyGlnPheLeuValAlaSnIleProProAlaValSerAspSerAspSerPro	299
QY	997	-----CGGAGCGGTGGCAGAGCGC-----	1017
Db	300	LeuGluAspIleSerAspLeuAlaSerGluLeuSerGlnAlaProProSerHisIlePro	319
QY	1018	---CGGCAGCGCTGAGCTGACGGCGCAGGAGTCTCTCATCGAACGCGCTGGCATGGAG	1074
Db	320	ProProArgHisAlaGlnArgSerProGluAlaSerGlnThrAspSerProValGlu	339
QY	1075	TCCACACAGATCTCCAGGAGCAGCAGGAGATGGAGCGCGCAAGGAGAAATATGCC	1134

Db	340	SerProArgLeu	-----ArgArgGluSerSerValAsp	350
Qy	1135	CAGAGGCACGACGAGGAATAGACATACCGAAGGAGATGGACAGATTCTAGAGGAG	1194	
Db	351	SerArgThrIleSerGluProAspGluInArgSerGlu	-----363	
Qy	1195	GAAGAGAAGTTTAAGAACCAATGGGAAGAAGACTGGGGCTCAAAGGAACAGCTACTCTTG	1254	
Db	364	-----Leu	364	
Qy	1255	CCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATGATCAG	1314	
Db	365	ProArgGluSerTyrAspIleTyrArgValProSerSerGlnSerMetGluAspArg	384	
Qy	1315	GGAGTGGAACTGAGCTCGAGCCCGCAGATGACCTGGATGGAGGCACGGAGGACGAGGA	1374	
Db	385	GlyTyrSerPro	-----388	
Qy	1375	GAGCAGGATTTCGGGAATATAGGAAGGCTTTGACCCCTACTCTATGTTCAACCCACAG	1434	
Db	388	-----388		
Qy	1435	CAGATCATGGGAGGATGTCGGCTCCTACGCATCAAGAAGGAGGATCCTTAGACCTG	1494	
Db	389	-----AspThrArgValValArgPheLeuLysGlyLysSerIleGlyLeu	403	
Qy	1495	CCCTGGAAGCGGTGTGGACTCCCCATTGGGAAGGTGGTTCGTCTGCTGTATGAG	1554	
Db	404	ArgLeuAaGlyGlyAsnAsp	-----ValGly---IlePheValSerGlyVal---Gln	419
Qy	1555	CGGGAGACTGCTGAGCGGCATGGTGGCATTTGTAAGGGGAGAGATCATGGCAATCAAC	1614	
Db	420	AlaGlySerProAlaaspGlyInGlyIleGlnGluGlyAspGlnIleLeuGlnValAsn	439	
Qy	1615	GGCAAGATTGTACAGACTACACCTGGCTGGCTGAGGCTGACGCTGCCCTCGACAAGCGCTGG	1674	
Db	440	AspValProPheGlnAsnLeuThrArgGluGluAa	-----451	
Qy	1675	AATCAGGCGGGAGCTGGATCGACACTTGTGGTTGGCTGCTGCCGCCCAAGAGAGATGAC	1734	
Db	452	-----ValGlnPheLeuLeuGlyLeuProGlyGluGluMetGlu	465	
Qy	1735	-----GATGAGCTGACCTCTTTGCTGAAGTCCAAAAGGGGAAC	1773	
Db	466	LeuValThrGlnArgLysGlnAspIlePheTyrLysMetValGlnSerArgValGlyAsp	485	
Qy	1774	CAAAATTCACGGTTAGAAACAGTAGCTCCGGCCCCACCTCGTGAACACAAAGCCTCGG	1833	
Db	486	SerPheTyrIleArgThrHisPheGluLeuGluPro	-----SerProPro	500
Qy	1834	ACCAAGCCTT-----CAGAGAGCCACATGACACACACC-----AGNTGGCATCCTTGG	1881	
Db	501	SerGlyLeuGlyPheThrArgGlyAspValPheHisValLeuAspThrLeuHisPro	519	
Qy	1882	GACCTGAATCTATCACCCAGGAATCTCAAACCTCCCTTTGGCCCTGAACCCAGGCGCAGATA	1941	
Db	520	-----GlyProGlyGlnSer	524	
Qy	1942	AGGAACAGCTCGGGCCACTTTTTTGAAGGCCAATGTGGAGAAAGGAGGACGACCGCGGT	2001	
Db	525	HisAlaArgGlyGlyHisTrpLeu	-----AlaValArg	535
Qy	2002	TTGGGAGAGATCTCAAGGATCCAGACTCTCATTCCTTTCT	2043	
Db	536	MetGlyArgAspLeuLeuArgGluGlnLeuArgGlyIleLeuPro	549	

RESULT 14

AAV53753

ID AAY53753 standard; Protein; 2037 AA.

XX

AC AAY53753;

XX
DT 22-FEB-2000 (first entry)


```

Db 1410 ProGlyAsnAlaValGluProLeuProSerAsnSerGluAsnLeuGlnAsnLysGluThr 1429
QY 1456 CGGCTCCTACGCATCAAGAGGAGGATCCTAGACCTGGCCCTG----- 1500
Db 1430 GluProThrValThrSerAspAlaAlaValAspLeuSerPheLysAsnValGln 1449
QY 1501 -----GAAGGGGTCTG-----GACTCCCCC 1521
Db 1450 HisLeuGluLeuProLysAspGlnGlyGlyLeuGlyLeuAlaLeuSerGluGluAspThr 1469
QY 1522 ATGGGAGGGTGGCTTCGTGTGTATGAGCGGGAGCTGCTGAGCGCATGGTGGC 1581
Db 1470 LeuSerGlyValIleIleLysSerLeuThrGluHisGlyValAlaAlaThrAspGlyArg 1489
QY 1582 ATTTGGAAGGGGAGAGATCGCATCAACGGCAGATTTGACAGACTACACCTG 1641
Db 1490 LeuLysValGlyAspGlnIleLeuAlaValAspAspGluIleValValGlyTyrProIle 1509
QY 1642 GCTGAGGCTGACGCTGCCCTGCAGAGCCCTGGAATCAGGCGGGGACTGGATGCACCTT 1701
Db 1510 GluLysPheIleSerLeuLeuLysThrAla----- 1519
QY 1702 GTGTTGCCGTCTGCCCCCAAGAGGATATGACGATGAGTCTTCTTCTGCTGAA-GTC 1760
Db 1520 -----LysMetThrValLysLeuThrIle 1527
QY 1761 CAAAGGGGAAACCAATTACGCGTTAGGNAACAGTGAAGCTCGGCGCCACCTCGTGAA 1820
Db 1528 HisAlaGluAsnProAspSerGlnAlaValProSerAlaAlaGlyAlaAlaSerGlyGlu 1547
QY 1821 CACAAAGCCTCGGACCGCTTGAGAGAGCCACATGACACACACAGATGGCATCCTTG 1880
Db 1548 LysLysAsnSerSerGlnSerLeuMetValPro-----GlnSerGlySerPro 1563
QY 1881 GGACCTGAATCTATC-----ACCCAGGAATCTCAAACTCCC-----TTTGGCCTGAA 1928
Db 1564 GluProGluSerIleArgAsnThrSerArgSerThrProAlaIlePheAlaSerAsp 1583
QY 1929 CCA----- 1931
Db 1584 ProAlaThrCysProIleIleProGlyCysGluThrThrIleGluIleSerLysGlyArg 1603
QY 1932 -----GGCCAGATAAGGAACAGCTCGGCGCCACTTTTGA 1967
Db 1604 ThrGlyLeuGlyLeuSerIleValGlySerAspThrLeuLeuGlyAlaIleIle 1623
QY 1968 AGGCCAATGTGGAGGAAGGAGGAGCAGCCGCTTTGGGAGAAGATCTCAAGGATCCAGA 2027
Db 1624 HisGluValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaGlyAspGln 1643
QY 2028 CTC 2030
Db 1644 Ile 1644

RESULT 15
AAU87918
ID AAU87918 standard; Protein; 1111 AA.
XX
AC AAU87918;
XX
DE
XX
XX 05-JUN-2002 (first entry)
XX Human PDZP1 protein.
XX Human; PDZ domain; WW domain; rat; cow; mouse; fruitfly; protein therapy;
XX gene therapy; PDZ-mediated disease; inward potassium channel; WBP;
XX dimer inhibitor peptide; carboxylate binding loop.
XX Homo sapiens.
XX
XX WO200207751-A1.
XX
XX 31-JAN-2002.
PD

```

```

XX
PF 24-JUL-2001; 2001WO-US23269.
XX
PR 25-JUL-2000; 2000US-221215P.
PR 28-NOV-2000; 2000US-0723810.
XX
PA (AXCE-) AXCELL BIOSCIENCES CORP.
XX
PI Herrero J, Pirozzi G, Uveges A;
XX
DR WPI; 2002-195842/25.
XX
PT Methods for identifying polypeptides comprising PDZ domains, the
PT polypeptides and their encoding nucleic acids, useful for the diagnosis
PT and treatment of PDZ related disorders.
XX
PS Claim 43; Fig 8; 225pp; English.
XX
CC The invention relates to methods for identifying polypeptides comprising
CC PDZ domains, and their encoding nucleic acids. The sequences are used to
CC identify modulators of their expression, function and activity, for use
CC in the diagnosis and treatment of PDZ related disorders. Antibodies
CC against the proteins and cells that produce them may be used for the
CC treatment of PDZ-mediated disease states. Sequences AAU87918-4
CC represent proteins containing PDZ domains, fragments of these proteins
CC and other related peptides used in the methods of the invention.
XX
SQ Sequence 1111 AA;

```

Alignment Scores:

```

Pred. No.: 1,76e-13 Length: 1111
Score: 286.50 Matches: 140
Percent Similarity: 35.41% Conservative: 110
Best Local Similarity: 19.83% Mismatches: 219
Query Match: 6.47% Indels: 238
DB: 23 Gaps: 24

```

US-09-502-945-4 (1-2236) x AAU87918 (1-1111)

```

QY 388 CTGGCGCTGAGTGGTGGTGGCTGGAGTTGGTGGTGGCTCTTCATCTCCACCTC 447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 LeuGlyMetThrValSerAlaAsnLysAsp---GlyLeuGlyMetIleValArgSerIle 82
QY 448 ATCAAGAGCGGTGAGCAGACAGCGTCCGG---CTCAGGTAGGGACGAGATCGTCCGG 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 83 IleHisGlyGlyAlaIleSerArgAspGlyArgIleAlaIleGlyAspCysIleLeuSer 102
QY 505 ATCAATGGATATTCATCTCTCTGTCACCATGAGGAGGTGATCAACCTCATTCGAACC 564
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 103 IleAsnGluGluSerThrIleSerValThrAsnAlaGlnAlaArgAlaMetLeuArg--- 121
QY 565 AAGAAAACCTGTGTCCATCAAGTAGACACATCGGCTGATCCCTGTAAGAACTCTCT 624
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 -----ArgHis---SerLeulle-----GlyPro 128
QY 625 GATGAGCCCTCACTTGG-----CAGTATGTGGATCACTTTGTGCGAATCTGGG 675
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 129 AspIleLysIleThrTyrValProAlaGluHisLeuGluGluPheLysIleSerLeuGly 148
QY 676 GCGCGCGAGGAGCGCTG----- 693
Db 149 GlnGlnSerGlyArgValMetalLeuAspIlePheSerSerTyrThrGlyArgAspIle 168
QY 694 -----GGCTCCCTCGAATCGGAAACCAAGGAG----- 723
Db 169 ProGluLeuProGluArgGluGluGlyGluGluSerGluLeuGlnAsnThrAla 188
QY 724 -----AAGAAGGTC 732
Db 189 TyrSerAsnTrpAsnGlnProArgValGluLeuTrpArgGluProSerLysSerLeu 208
QY 733 TTCATCAGCCTGGTAGGCTCCGAGGCTTGGCTGCACGATTTCCAGCGGCCCATTCAG 792
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 209 GlyIleSerIleValGlyGlyArgGlyMetGlySerArgLeuSerAsnGlyGluValMet 228
QY 793 AAGCTGCTCATCTTTATGAGCAGCTGAACCTGGCTCCCTGCTGCTGAGAGTGGGA--- 849
Db 229 Arg---GlyIlePheIleHisValLeuGluAspSerProAlaGlyLysAsnGlyThr 247
QY 850 TTGGAGATAGGGGACAGATTGTGGAAGTCAATGGCTGCTGACTTCTCTTAACCTGGATCAC 909
Db 248 LeuLysProGlyAspArgIleValGluValAspGlyMetAspLeuArgAspAlaSerHis 267
QY 910 AAGGAGGCTGTAAATGCTGCTGAAATAGCCGACGCTGACCATCTCCATTGCTAGCTGCA 969
Db 268 GluGlnAlaValGluAlaIleArgLysAlaGlyAsnProValValPheMetValGlnSer 287
QY 970 GCTGCGGGGAGCTGTTATGACACACGGGAGCGGCTGGCAGAGCGCGGCGCAGCTGAG 1029
Db 288 -----IleIleAsnArgProArgAlaProSerGlnSerGluSerGlu 301
QY 1030 CTGACGCGGAGGAGCTTCTCATGTCAGAGCGGCTGGCGATGGAGTCCAAACAAGATCCTC 1089
Db 302 ProGluLysAlaProLeuCys----- 308
QY 1090 CAGGAGCAGCAGGAGATGGAGCGGCAAGGAGAGAAAGAAATTGCCAGAGGCGACAGAG 1149
Db 309 -----SerValProProProProSerAlaPheAlaGluMetGlySerAsp 324
QY 1150 GAAATATGAGATACCGGAAGAGATGGAACAGATTGTAGAGGAGGAAGAGATTAAAG 1209
Db 325 HisThrGlnSerSerAlaSerLysIleSerGlnAspValAspLysGluAspGluPheGly 344
QY 1210 AAGCAATGGGAA-----GAAGACTGGGGCTCAAGAGAACACTACTCTTGGCTAAA 1260
Db 345 TyrSerTrpLysAsnIleArgGluArgTyrGly----- 355
QY 1261 ACCATCACTGCTGAGTACACCCACTACCCCTCCGACAG----- 1299
Db 356 ThrLeuThrGlyGluLeuHisMetIleGluLeuGluLysGlyHisSerGlyLeuGlyLeu 375
QY 1300 -----CCA 1302
Db 376 SerLeuAlaGlyAsnLysAspArgSerArgMetSerValPheIleValGlyIleAspPro 395
QY 1303 AAGTATGATCAGGGAGTGGAACCTGAGCTCGAGCCCGCAGACTGCTGGATGGAGGCAGG 1362
Db 396 AsnGlyAlaAlaGlyLysAspGlyArgLeuGlnIleAlaAspGluLeu-----Leu 412
QY 1363 GAGGAGCAGGAGCAGCAGGATTTCGGAATATAGAGAAAGGCTTTGACCCCTACTCTATG 1422
Db 413 GluIleAsnGlyGlnIleLeuTyrGlyArgSerHisGlnAsnAlaSerSerIleIleLys 432
QY 1423 TTCACCCCGACAGACATC----- 1440
Db 433 CysAlaProSerLysValLysIleIlePheIleArgAsnLysAspAlaValAsnGlnMet 452
QY 1440 -----ATGGGGAAG 1449
Db 473 LysGluThrGluProThrValThrThrSerAspAlaAlaValAspLeuSerSerPheLys 492
QY 1450 GATGTCGGCTCCTACGCATC---AAGAAGAGAGGATCCTTAGACCTGGCCCTGGGAAGGC 1506
Db 493 AsnValGlnHisLeuGluLeuProLysAspGlnGlyGlyLeuGlyIleAlaIle----- 510
QY 1507 GGTGTGGACTCCCCCATNTGGGAAGGTGCTGTTCTGTGTATGATGAGCGGGGAGCTGCT 1566
Db 511 SerGluGluAspThrSerGlyValIleIleLysSerLeuThrGluHisGlyValAla 530
QY 1567 GACCGCATGGTGGCATTTGTAAAGGGGAGCAGATCATCGCAATCAACGGCAAGATTGTG 1626
Db 531 AlaThrAspGlyArgLeuLysValGlyAspGlnIleLeuAlaValAspGluIleVal 550

QY 1627 ACAGACTACACCTGGCTGAGCTGAGCTGCCCTGCAGGAAGGCCTGCAATCAGGGCGGG 1686
Db 551 ValGlyTyrProIleHisAlaGluAsnProAspSerGlnAlaValProSerAlaAlaGly 585
QY 1687 GACTGGATCAGCTTGTGGTGGCTGCTGCCCTGCCCCCAAGAGTATGACGATGAGCTGACC 1746
Db 566 -----LysMetThr 568
QY 1747 TTCTTGTCTGAA-GTCCAAAAGGGGAAACCAAAATTCACGCGTTAGGAAACAGTGAAGCTCCG 1805
Db 569 ValLysLeuThrIleHisAlaGluAsnProAspSerGlnAlaValProSerAlaAlaGly 588
QY 1806 GCCCACCCTCGTGAACACAAAGCCTCGACACAGCCTTGAGAGAGGCCACATGACACACAC 1865
Db 589 AlaAlaSerGlyGlyLysLysAsnSerSerGlnSerLeuMetValPro----- 604
QY 1866 CAGATGCGCATCCTTGGGACCTGAATCTATC-----ACCCAGGAATCTCAAACTCCC- 1916
Db 605 GlnSerGlySerProGluProGluSerIleArgAsnThrSerArgSerThrProAla 624
QY 1917 ---TTTGGCCCTGAACCA----- 1931
Db 625 IlePheAlaSerAspProAlaThrCysProIleIleProGlyCysGluThrThrIleGlu 644
QY 1932 -----GGGCCAGATAAGGAACAGCTC 1952
Db 645 IleSerLysGlyArgThrGlyLeuGlyLeuSerIleValGlyGlySerAspThrLeuLeu 664
QY 1953 GGGCCACTTTTGAAGGCCAATGTGGAGGAAGGGAGCAGCCGCTTTGGGAGAAGA 2012
Db 665 GlyAlaIleIleIleHisGluValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeu 684
QY 2013 TCTCAAGGATCCAGACTC 2030
Db 685 TrpAlaGlyAspGlnIle 690

Search completed: March 21, 2003, 12:43:57
Job time : 77.6146 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:34:32 ; Search time 17.3222 Seconds
(without alignments)
7595.986 Million cell updates/sec

Title: US-09-502-945-4
Perfect score: 4120
Sequence: 1 cctggccggtcggtgcg.....tattttccagcttaaaaaa 2236

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+np.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09502945/runat_14032003_101059_19145/app_query.fasta_1.10979
-DB=Issued_Patents_AA -QFWT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_80.erna_14032003_101059_19145 -NCPU=6 -ICPU=3
-NO_XLPX -NO_MWAP -LARGOQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	267.5	6.5	2037	US-09-306-998-3	Sequence 3, Appli
2	259.5	6.3	1881	US-09-233-086-3	Sequence 3, Appli
3	238.5	6.3	724	US-09-562-737-21	Sequence 21, Appli
4	242.5	5.9	724	US-09-562-737-28	Sequence 28, Appli
5	228.5	5.5	724	US-09-562-737-23	Sequence 23, Appli
6	220.5	5.4	724	US-09-562-737-26	Sequence 26, Appli
7	215.5	5.2	724	US-09-562-737-24	Sequence 24, Appli
8	201.5	4.9	1112	US-09-045-632-2	Sequence 2, Appli
9	201.5	4.9	1112	US-09-045-632-3	Sequence 3, Appli
10	201	4.9	724	US-09-562-737-30	Sequence 30, Appli
11	200.5	4.9	604	US-09-045-632-14	Sequence 14, Appli
12	200.5	4.9	702	US-09-045-632-15	Sequence 15, Appli

13	200.5	4.9	1018	3	US-09-045-632-16	Sequence 16, Appli
14	200.5	4.9	1061	3	US-09-045-632-32	Sequence 32, Appli
15	199.5	4.8	724	4	US-09-562-737-22	Sequence 22, Appli
16	195.5	4.7	2485	4	US-09-290-640-46	Sequence 46, Appli
17	195	4.7	450	2	US-08-665-037-2	Sequence 2, Appli
18	195	4.7	450	2	US-08-666-087-2	Sequence 2, Appli
19	195	4.7	450	2	US-08-732-870-2	Sequence 2, Appli
20	195	4.7	1050	3	US-09-045-632-50	Sequence 50, Appli
21	192	4.7	1050	3	US-09-045-632-49	Sequence 49, Appli
22	191.5	4.6	724	4	US-09-562-737-27	Sequence 27, Appli
23	191.5	4.6	2466	3	US-09-080-855-12	Sequence 12, Appli
24	191.5	4.6	2466	5	PCT-US94-09943-2	Sequence 2, Appli
25	190.5	4.6	724	4	US-09-562-737-25	Sequence 25, Appli
26	190	4.6	2465	2	US-08-596-291-3	Sequence 3, Appli
27	190	4.6	2465	3	US-09-100-804-3	Sequence 3, Appli
28	187	4.5	960	4	US-09-219-849-6	Sequence 6, Appli
29	186	4.5	610	1	US-08-410-804-1	Sequence 1, Appli
30	186	4.5	610	1	US-08-259-514-1	Sequence 1, Appli
31	186	4.5	610	2	US-08-858-311-1	Sequence 1, Appli
32	185.5	4.5	507	3	US-09-045-632-13	Sequence 13, Appli
33	184	4.5	724	4	US-09-562-737-29	Sequence 29, Appli
34	182	4.4	519	3	US-08-997-445D-2	Sequence 2, Appli
35	181.5	4.4	552	4	US-09-219-849-7	Sequence 7, Appli
36	177.5	4.4	552	4	US-09-219-849-7	Sequence 7, Appli
37	177	4.3	1461	4	US-09-585-887-9	Sequence 9, Appli
38	177	4.3	1461	4	US-09-289-578-9	Sequence 9, Appli
39	173.5	4.2	822	4	US-09-219-849-49	Sequence 49, Appli
40	172.5	4.2	1341	3	US-08-963-825-18	Sequence 18, Appli
41	172.5	4.2	1341	4	US-09-500-811-18	Sequence 18, Appli
42	172.5	4.2	1341	4	US-09-570-573-18	Sequence 18, Appli
43	172.5	4.2	1341	4	US-09-548-608-18	Sequence 18, Appli
44	172	4.3	960	4	US-09-219-849-6	Sequence 6, Appli
45	170.5	4.1	1078	3	US-08-963-825-21	Sequence 21, Appli

ALIGNMENTS

RESULT 1
US-09-306-998-3
; Sequence 3, Application US/09306998
; Patent No. 6291173
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtighian, Sean V.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/306,998
; CURRENT FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 60/084,740
; EARLIER FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-306-998-3

Alignment Scores:				
Pred. No.:	8 11e-14	Length:	2037	
Score:	267.50	Matches:	142	
Percent Similarity:	35.81%	Conservative:	109	
Best Local Similarity:	20.26%	Mismatches:	223	
Query Match:	6.49%	Indels:	228	
DB:	4	Gaps:	26	

US-09-502-945-4 (1-2236) x US-09-306-998-3 (1-2037)

Qy	388	CTCGGCTGAGTGTGCTGGTGGCTGGAGTTTGGCTGTTCATCTCCACCTC	447
		:	
		::: :::	
Db	1018	LeuclyMetThrValSerAlaAsnLysAsp---GlyLeuGlyMetIleValArgSerIle	1036
Qy	448	ATCAAGCGCGGTACAGGACAGACGTCGGG---CTCAGGTAGGGACAGATGTCGCG	504

RESULT 2
US-09-233-086-3
; Sequence 3, Application US/09233086
; Patent No. 6337192
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtligian, Sean V.

```

; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: MMSC1 - An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC1 Gene
; CURRENT APPLICATION NUMBER: US/09/233,086
; CURRENT FILING DATE: 1999-01-19
; EARLIER APPLICATION NUMBER: US 60/071,861
; EARLIER FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 3
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-233-086-3

```

```

Alignment Scores:
Pred. No.: 3,65e-13 Length: 1881
Score: 259.50 Matches: 149
Percent Similarity: 37.03% Conservative: 115
Best Local Similarity: 20.90% Mismatches: 232
Query Match: 6.30% Indels: 217
DB: 4 Gaps: 31

```

US-09-502-945-4 (1-2236) x US-09-233-086-3 (1-1881)

```

QY 85 CQTGGCCAGCCATGGACCGAAGAGTGGCCCGAGATCCCGCATAGGTGATTTCTG 144
DB 169 ProGlySerValAlaAspArgAsp-----GlnArgLeu 179

QY 145 ATTGAATATGATGCAGAGAGGACTATCTATGATGTGCTGCGAATGTACACAGACC 204
DB 180 LysGluAsnAsp-----GlnIleLeuAlaIleAsnHisThrPro 192

QY 205 ATGGAC-----GTGGCGGTGCTC-----GTGGGAGAC 231
DB 193 LeuAspGlnAsnIleSerHisGlnGlnAlaLeuLeuGlnGlnThrThrGlySer 212

QY 232 CTGAAGCTGGTCATC-----AATGAACCC-----AGCGGTGCTGCTG 270
DB 213 LeuArgLeuIleValAlaArgIleProValHisThrLysSerSerSerSerSerLeu 232

QY 271 TTTGATGCATTTCGGCCGCTGATCCCACTGAAGCACCAGGTGGATATGATCAGTGACC 330
DB 233 AsnAspThrThrLeuProGluThrValCysTrpGlyHisValGluGluValGluLeu 252

QY 331 CCCCAGCGCTCCAGGAAGCTGAAGAGGTGCTCTGGACCGTCTCCACCCCGAAGGCTC 390
DB 253 AsnAspGlySer-----GlyLeu 258

QY 391 GGCCTGAGTGTGCGTGGTGGCTGGAGTTTGCTGTGGCTCTTCATCTCCACCTCATC 450
DB 259 GlyPheGlyIleValGlyGlyLysThr-----SerGlyValValValArgThrIleVal 276

QY 451 AAAGCGGTGACGACAGACAGCGTCCGG-----CTCCAGGTAGGGAGCAGATCGTCCGGATC 507
DB 277 ProGlyGlyLeuAlaAspArgAspGlyArgLeuGlnThrGlyAspHisIleLeuLysIle 296

QY 508 AATGATATTCATCTCCTCTGTAACCATGAGGAGGTATCAACCTCATTCGAACC--- 564
DB 297 GlyGlyThrAsnValGlnGlyMetThrSerGluGlnValAlaGlnValLeuArgAsnCys 316

QY 565 AAGAAACTGTCTCATCAAGTGAAGACATCGCGCTGATCCCGTGAAGAGTCTCCT 624
DB 317 GlyAsnSerValArgMetLeuValAlaArgAspProAlaGlyAspIleSerValThrPro 336

QY 625 GATGAGCCCTCACCTTGCAGTATGTGATCATGTTGTGTCGGAATCGGGGGCGTCCGA 684
DB 337 ProAlaProAlaAlaLeuProValAlaLeuProThrValAlaSerLysGly----- 353

QY 685 GCGAGCTGGGCTCCCTGGAAATCGGGAAACAAGGAGAGAGGTCTTCATCAGGCTG 744
DB 354 -----ProGlySerAspSerSerLeuPheGluThrTyrAsnValGluLeu 368

```

```

QY 745 GTA-----GGTCCGAGCGCCTTGCTCCAGCATTTCCAGCGCG 783
DB 369 ValArgLysAspGlyGlnSerLeuGlyIleArgIleValGlyThrValGlyThrSerHis 388

QY 784 CCCATCCAGAACCTGGCCTTTATCAGCCATGTGAACCTGCTCCCTGCTGCTGAG 843
DB 389 ThrGlyGluAlaSerGlyIleThrValLysSerValIleProGlySerAlaAlaThrHis 408

QY 844 GTGGGA---TTGGAGATAGGGACAGATTGTCGAAGTCAATGGCTGCGACTTCTCTAAC 900
DB 409 AsnGlyHisIleGlnValAsnAspLysIleValAlaValAspGlyValAsnIleGlnGly 428

QY 901 CTGATCATCAAGAGGCTGTAAATGTCTGCTGAAATAGCCGACGCTGACCATCTCCATT 960
DB 429 PheAlaAsnHisAspValValGluValLeuArgAsn----- 440

QY 961 GTAGCTGCAGCTGCGCGGAGCTGTTTCATGACA----- 993
DB 441 -----AlaGlyGlnValValHisLeuThrLeuValArgArgLysThrSerSerSer 457

QY 994 -----GACCGGAGCGCTGGCAGAGCGCGCGAGCGTGTGAG 1029
DB 458 ThrSerProLeuGluProProSerAspArgGlyThrValValGlu-----Pro 473

QY 1030 CTGAGCGGAGGAGCTTCTCATGACAGAGCGCTGGCGATGGAGTCCACACAGATCCTC 1089
DB 474 LeuLysProProAlaLeuPheLeuThrGlyValValGluThrGluThrAsn---ValAsp 492

QY 1090 CAGGACGAGGAGGATGGCGCAAGAGAGAAAGAAATTCGCCAGAGGAGCAGAG 1149
DB 493 GlyGluAspGluGluIle-----LysGluArgIleAspThrLeuLysAsn 507

QY 1150 GAAATGAGAGATACCGGAAGAGAGATGGAACAGATTGTAGAGAG-----GAAGAGAAGTTT 1206
DB 508 AspAsnIleGlnAlaLeuGluLysLeuLysValProAspSerProGluAsnGluLeu 527

QY 1207 AAGAACAAATGGGAAGAGACTGGGGCTCAAGAGACAGCTACTCTTGGCTTAAACCATC 1266
DB 528 LysSerArgTrpGluAsnLeuLeuGlyProAspTrpGluValMetValAlaThrLeuAsp 547

QY 1267 ACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCAAGATGATGATCAG----- 1314
DB 548 ThrGlnIleAlaAspAspAlaGluLeuGln-----LysThrSerLysLeuLeuProIle 565

QY 1315 -----GGAGTGAACCTGAG----- 1329
DB 566 HisThrLeuArgLeuGlyValGluValAspSerPheAspGlyHisHisThrIleSerSer 585

QY 1330 -----CTCAGAGCCCGCAGATGACCTG--- 1350
DB 586 IleValSerGlyGlyProValAspThrLeuGlyLeuLeuGlnProGluAspGluLeuLeu 605

QY 1351 -----GATGGA----- 1356
DB 606 GluValAsnGlyMetGlnLeuThrGlyLysSerArgArgGluAlaValSerPheLeuLys 625

QY 1356 ----- 1356
DB 626 GluValProProPheThrLeuValCysCysArgArgLeuPheAspAspGluAlaSer 645

QY 1357 -----GGCAGCGAGGAGCAGGAGGAGCAG 1380
DB 646 ValAspGluProArgThrGluThrSerLeuProGluThrGluValAspHisAsnMet 665

QY 1381 GATTTCGGAATATGAGGAGGCTTTGACCCCTACTCTATGTTTACCCCGCAGAGCATC 1440
DB 666 AspValAsnThrGluGluAspAspGlyGluLeuAlaLeuTrpSerProGlu----- 683

QY 1441 ATGGGGAAGGATGTCCGCTCTCTACGATCAAGAAGAG-----GGATCCTTA 1488
DB 684 -----ValLysIleValGluLeuValLysAspCysGlyLeuGlyPheSer 699

QY 1489 GACCTGGCCTGGAAAGCGGTGTGGACTCCCCATTTGGGAAGGTGCTGCTGCTGTG 1548

```

```

Db 700 IleLeuAspTyrGlnAspProLeuAspProThrArgSerValIleValIleArgSerLeu 719
    |||    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1549 TATGAGCGGGAGCTCTCAGCGGCATGTGGCATTTGAAAGGGGACGAGATCATGGCA 1608
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 720 ValAlaAspGlyValAlaGluArgSerGlyGlyLeuLeuProGlyAspArgLeuValSer 739
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1609 ATCAAGCGCAAGATTGTGACAGACTACACCCTGGCTGAGGCTGACGCTCCCTGCAGAAG 1668
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 740 ValAsnGluTyrCysLeuAspAsnThrSerLeuAlaGluAla--ValGluIleLeuLys 758
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1669 GCCTGGAATCAGCGGGGAGCTGGATCGACCTTGTG---GTTGCCCTCTGCCCCCA--- 1722
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 759 AlaValProGly-----LeuValHisLeuGlyIleCysLysProLeu 773
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1723 ---AAGGAGTATGACGATGACCTGACCTTCTCTGCTG-----AAG 1758
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 774 ValGluAspAsnGluGluSerCysTyrIleLeuHisSerSerSerAsnGluAspLys 793
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1759 TCCAAAGGGGAACCAAAATTCACGGGTAGGAAACAGT 1797
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 794 ThrGluPheSerGlyThrIleHisAspIleAsnSerSer 806
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||

RESULT 3
US-09-562-737-21
; Sequence 21, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 21
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-21

Alignment Scores:
Pred. No.: 2,63e-13 Length: 724
Score: 256.50 Matches: 127
Percent Similarity: 39.74% Conservative: 90
Best Local Similarity: 23.26% Mismatches: 200
Query Match: 6.27% Indels: 129
DB: 4 Gaps: 25

US-09-502-945-4 (1-2236) x US-09-562-737-21 (1-724)
QY 340 TCCAGGAAGCTGAAGAGGTGCGTCTCGACCGCTGCACCCCGAAGCGCTCGGCCTGAGT 399
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 155 AlaGluLysIleIleGluIleLysLysLys---GlyProLysGlyLeuGlyPheSer 173
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 400 GTGCGTGGTGGC-----CTGGAGTTTGGCTGTGGGCTCTTCATCTCCAC 444
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 174 IleAlaGlyValGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 193
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 445 CTCATCAAGCGGTGACGAGCAGACAGCGTCGGG---CTCCAGGTAGGGGACGAGATCGTC 501
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 194 IleIleGluGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 213
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 502 CGGTAATGATATTCATCTCTCTCTGACCCATGAGGAGGTTCATCAACCTCATTCGA 561
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 214 AlaValAsnSerValGlyLeuGluAspValMethHisGluAspAlaValAlaLeuLys 233
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 562 ---ACCAAGAAACTGTGTCCATCAAAAGTG-----AGACACATCGGCCTGATCCCC 609
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
```

```

Db 234 AsnThrTyrAspValValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 253
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 610 GTGAAAGCTCTCCTGATGAGCCCTCCTACTTGG---CAGTATGTGGATCAGTTTGTGTCG 666
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 254 SerTyrAlaProProAspIleThrSerTyrSerGlnHisLeuAspAsnGluIleSer 273
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 667 GAATCTGGGGCGTCGAGGACGCTGGC----- 696
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 274 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 289
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 697 -----TCCCTGGAAATCGGAA-----AACAG 720
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 290 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluGluAspIleProArg 309
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 721 GAGAAAGAGGTCTTCATCAGCTGAGCTCCCGAGCGCTTGGCTGCAGCATTTCCAGC 780
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 310 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 329
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 781 GCCCCATCCAGAACCTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 330 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 348
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 832 CTGCTCTGCTGAGGTGGATTGGAGATAGGGACACAGATTGTGGAAGTCAATGGCGTCGAC 891
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 349 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 366
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 892 TTCTCTAACTGGATCACAAAGAGGCTGTAAATGTCTGAAATAATAGCGCGCATGACC 951
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 367 LeuArgAsnAlaSerHisGluGlnAlaIleAlaLeuLysAsnAla--GlyGlnThr 385
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 952 ATCTCCATTGTAGTCGACGTGGCGGAGCTGTCATCAGACAGCGGGCGCGCTGGCA 1011
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 386 ValThrIleIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1012 GAGCGCGCAGCGCTGAGCTGCAGCGGAGGAGCTTTCATCAGACAGCGGGCGGTG 1065
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1066 -----GCGATGAGTCCAAAC----- 1080
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1080 ----- 1080
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheHisPheGlyAsp 458
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1081 -----AAGATCTCCAGGACGAGAGATGGAGCGCAAGAGAGAAAGAAATGGC 1134
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 459 ValLeuHisValIleAspAlaSerAspGluGluTyrTrpGlnAlaArgValHisSer 478
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1135 CAGAAGGCACGACAGGAGAAATGAGAGATACCGGAAGAGAGATGGACACAGATTGTAGAGGAG 1194
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 479 AspSerGluThrAsp-----AspIleGlyPheIleProSerLys 491
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1195 GAAGAGAAGTTTAAGAACGATGGGNA-----GAAGACTGGGCTCA----- 1236
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 492 ArgArgValGluArgGluTrpSerArgLeuLysAlaLysAspTrpGlySerSerSer 511
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1237 -----AAGAACAGCTACTCTTGCCTTAAACCATCACT---GCTGAGTGA 1278
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 512 GlySerGlnGlyArgGluAspSerValLeuSerTyrGluThrValThrGlnMetGluVal 531
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1279 CAC-----CCAGTACCCTTCGCAAGCCAAAGATGATCATCGAGGAGTGGAACTGAG 1329
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 532 HisTyrAlaArgProIleIleIleLeuGlyProThrLysAspArgAlaAsnAspLeu 551
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1330 CTCGAGCCCGAGATGACCTGGATGGAGCGACGAGGAGCAGGAGAGCAGGATTTCCGG 1389
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 552 LeuSerGluPheProAspLysPheGlySerCysValProHisThrThrArgProLysArg 571
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1390 AAATATGAGGAAGGCTTTCACCCCTACTCTATCTTCCACCCACAGCAGATCATGGGAAG 1449
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 572 GluTyrGluIleAspGlyArgAspTyrHisPheValSerSerArgGluLysMetGluLys 591
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
```

```
QY 1450 GATGTCGGCTCCTACCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAGAGCGGT 1509
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 592 AsplieglnAlaHisLysPheIleGluAlaGlyInTyrAsnSerHisLeuTyrGlyThr 611
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1510 GTGAGCTCCGCCATT-----GGGAAG-----GTGGTC 1536
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 612 SerValGlnSerValArgGluValAlaGluGlnGlyLysHisCysIleLeuAspValSer 631
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1537 GTTCTGCTGTATGAGCGGGAGCTGCTGAGCGGCATGGTGCAATGTG----- 1587
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 632 AlaAsnAlaValArgLeuGlnAlaAlaHisLeuHisProIleAlaIlePheIleArg 651
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1588 -----AAGGGGACGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTG 1641
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 652 ProArgSerLeuGluAsnValLeuGluIleAsnLysArgIleThrGluGlnAlaArg 671
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1642 GCTGAGCTGACGCTGCC 1659
||| |||
Db 672 LysAlaPheAspArgAla 677
||| |||

RESULT 4
US-09-562-737-28
; Sequence 28, Application US/09562377
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: US070708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 28
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-28

Alignment Scores:
Pred. No.: 5 8e-12 Length: 724
Score: 242.50 Matches: 142
Percent Similarity: 40.30% Conservative: 101
Best Local Similarity: 23.55% Mismatches: 227
Query Match: 5.89% Indels: 133
DB: 4 Gaps: 33

US-09-502-945-4 (1-2236) x US-09-562-737-28 (1-724)
QY 250 GAACCCAGCGCTGCTGCTGCTGTTGATGCCATTCCG-----CGCGTGATC 294
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 107 GlnGlnGlyArgLeuArgValAsnAspSerIleArgPheValAsnGluValAspValArg 126
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 295 CCAGTGAAGCACCAGGTGGATATGATCAGCTGACCCCC----- 333
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 127 GluSerThrHisSerAlaAlaValGluAlaLeuThrGluAlaGlySerIleValArgLeu 146
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 334 -----CGGCGC-----TCCAGGAGCTGAGGAGGTGCTCTGGACCGCTG 375
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 147 TyrTrpMetArgLysProProAlaGluLysTyrIleGluIleLysLeuLys--- 165
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 376 CACCCCAAGGCGCCGCTGAGTGTGCTGGTGGC-----CTGGAGTTT 420
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 166 GlyProAlaGlyLeuGlyPheSerIleAlaGlyGlyAspGlyAsnGlnHisIleProGly 185
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 421 GGCTGTGGCTTTCATCTCCCATCATCAAGCGGCTCAGGACAGCAGCGTCGGG--- 477
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 186 AspAsnGluIleTyrValThrLysIleIleGluCluGlyPheAlaAlaHisLysAspGlyArg 205
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 478 CTCACGAGTGGGACGAGATCGTCCGGATCAATGGATATTCATCTCTCTCTGTTACCCAT 537
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 206 LeuGlnGlyAspLysIleLeuAlaValAsnSerHisGlyLeuGluAspValMetHis 225
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 538 GAGGAGGTCAATCAACCTCTCGA---ACCAAGAAACCTGTGTCCATCAAAAGTGAGACAC 594
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 226 GluAspIleValAlaAlaLeuLysAsnThrTyrAspLysValTyrLeuLysValAlaLys 245
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 595 ATCGGCTGATCCCCGTGAAAGCTCT-----CTGTAGAGCCCTCACT-----TGG 642
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 246 ProSerLeuAlaTyrLeuSerAspSerTyrAlaProMetAspIleThrThrSerTyrSer 265
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 643 CAGTATGTGATCAGTTTGTGTCGGAATCGGGGCGTCGAGGCAGC----- 690
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 266 GlnHisAsnAspAsnGluIleSerHisSer---SerTyrGlnGlyThrAspTyrProThr 284
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 691 -----CTGGGCTCCCTCGGAATCGGAA----- 714
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 285 AlaMetThrArgThrSerProArgArgTyrSerProValSerLysAspLeuLeuGlyGlu 304
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 715 -----AACAGGAGAGAGAGTCTTCATCAGCCTGGTAGGCTCCGAGGCCTTGGC 765
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 305 GluAspIleThrArgGluProArgIleValIleHisValGlySerThrGlyLeuGly 324
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 766 TGCAGCATTTCCAGCGCCCCCATCCAGAGCCTGTCATCTTTATATCAGCCATGTG----- 819
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 325 PheAsnIleTrpGlyGlyGlu---AspGlyGluGlyIlePheTyrSerPheIleLeuAla 343
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 820 ---AAACCTGGCTGCTGTGTGAGTGGGATTCAGATAGGAGCAGCATGTTCGAA 876
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 344 GlyGlyProAlaAlaLeuSerGlyGlu-----LeuArgLysGlyAspIleLeuSer 361
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 877 GTCAATGGGTGCGATCTCTTAACCTGGATCAACAGGAGCGTCTAAATGTGCTGAAAAAT 936
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 362 ValAsnGlyValAspLeuGluAsnAlaSerHisGluGlnAlaAlaIlePheLeuLysAsn 381
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 937 AGCCGCGAGCTGACCATCTCCATTGTAGCTGCAGCTGGCGGAGCTGTTTCATGACAGAC 996
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 382 Ala---GlyGlnThrValThrGlyIleAlaGln-----TyrLysProGlu 395
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 997 CGGAGCGGCTGGCAGAGCGCGGCGAGCTGAGCTGCAGCGCAGGAGCTTCTCATGAC 1056
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 396 GluTyrHisArgPheGluAlaLysIleHisAspLeuIleGluGln-----LeuMetAsn 413
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1057 AAGCGGCTG-----GCGATGGAGTCCAAAGATCTCTCAG----- 1092
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 414 SerSerLeuGlyLysGlyThrAlaSerLeuArgSerAsnProLeuArgGlyPheTyrIle 433
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1093 -----GAGCAGCAGGAGATGGAG 1110
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 434 ArgAlaLeuPheMetTyrAspLysThrLysAspCysGlyPheAsnSerGlnAlaLeuSer 453
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1111 CGCAAGAGGAGAAAGAAATTCGCCAGAGGCGCAGAGGAGAAATGAGATACCGG--- 1167
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 454 PheHisPheGlyGlnValLeuHisValIleAspAlaSerAspArgGluTrpTrpGlnAla 473
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1168 -----AAGGAGATGGACAGATT-----GTAGAGGAGGAGAGAGAG 1203
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 474 ArgArgValHisThrAspSerGluThrAspAspIleGlyPheValProSerLysArgArg 493
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1204 TTTAAGAACCAATGG-----GAAGAGACTGG----- 1230
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 494 ValGluArgArgTrpTrpSerArgLeuLysAlaLysAspTrpTyrSerSerGlySer 513
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1231 ---GGGTCAAGGAGCAGACTTCTTCCTTAAACCATCACT---GCTGAGGTACAC--- 1281
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 514 GlnGlyArgGluAlaSerValLeuSerTyrGluThrValThrAspMetGluValHisTyr 533
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1282 -----CCAGTACCCCTTCGCAAGCCAAAGTATGATCAGGAGGTGGAACCTGAGCTCGAG 1335
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 534 AlaArgProIleGluIleLeuGlyProThrLysAspArgAlaPheAspLeuLeuSer 553
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1336 CCCGCAAGTACCTGGATGGAGGACGAGGAGGAGGAGGAGGATTTCCGGAAATAT 1395
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1450 GATGTCGGCTCCTACCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAGAGCGGT 1509
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 592 AsplieglnAlaHisLysPheIleGluAlaGlyInTyrAsnSerHisLeuTyrGlyThr 611
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1510 GTGAGCTCCGCCATT-----GGGAAG-----GTGGTC 1536
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 612 SerValGlnSerValArgGluValAlaGluGlnGlyLysHisCysIleLeuAspValSer 631
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1537 GTTCTGCTGTATGAGCGGGAGCTGCTGAGCGGCATGGTGCAATGTG----- 1587
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 632 AlaAsnAlaValArgLeuGlnAlaAlaHisLeuHisProIleAlaIlePheIleArg 651
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1588 -----AAGGGGACGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTG 1641
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 652 ProArgSerLeuGluAsnValLeuGluIleAsnLysArgIleThrGluGlnAlaArg 671
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1642 GCTGAGCTGACGCTGCC 1659
||| |||
Db 672 LysAlaPheAspArgAla 677
||| |||
```



```
Db 545 AspGluAlaAsnAspLeuSerGluPhePheAspLysPheGlySerCysValPro 564
|||... : : : ||| |||
QY 1369 CAGGAGAGCAGGATTCCGGAATATGAGGAAGCTTTGACCCCTACTCTATCTTCAAC 1428
||| : : : |||...||| |||
Db 565 HisGlyThrArgProLysArgGluTyGluHisGlyArgAspTyrHisPheValSer 584
||| : : : |||...||| |||
QY 1429 CCAGAGCAGATCATGGGAGGATGTCGGCTCTACGATCAAGAAGAGGGATCCCTTA 1488
||| : : : |||...||| |||
Db 585 SerIleGluLysMetGluLysAspIleAlaLysLysPheIleGluAlaGlyGluTyr 604
||| : : : |||...||| |||
QY 1489 GACCTGGCCCTGGAGGC--GGTGTGACTCC----- 1518
||| : : : |||...||| |||
Db 605 AsnLeuHisLeuTyrGlyThrSerValGlnSerMetArgGluValAlaGluGlnGlyLys 624
||| : : : |||...||| |||
QY 1519 ---CCCATTTGGGAAGGTGCTTCTCTGCTGTATGAGCGGGAGCTGTGAGCGGCAT 1575
||| : : : |||...||| |||
Db 625 HisAsnIleLeuAspValSerAlaAsnAlaValGlnArgLeuGlnAlaHisLeuHis 644
||| : : : |||...||| |||
QY 1576 GGTGCATTTGT-----AAGGGGACGAGATCATGGCAATCAACGGCAAG 1620
||| : : : |||...||| |||
Db 645 ProArgAlaIlePheIleArgProArgSerLeuSerAsnValLeuGluIleAsnLysArg 664
||| : : : |||...||| |||
QY 1621 ATTGTGACAGACTACACCTCGCTGAGCTGACGCTGCCCTGCAGAAAGGCTGGAATCAG 1680
||| : : : |||...||| |||
Db 665 IleValGluGluGlnAla-----ArgLysAlaPhe----- 674
||| : : : |||...||| |||
QY 1681 GCGCGGGAGCTGG 1692
||| : : : |||...||| |||
Db 675 -----AspTyr 676

RESULT 7
US-09-562-737-24
; Sequence 24, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gorthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562.737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-24

Alignment Scores:
Pred. No.: 1,07e-09 Length: 724
Score: 215.50 Matches: 110
Percent Similarity: 40.51% Conservative: 82
Best Local Similarity: 23.21% Mismatches: 179
Query Match: 5.23% Indels: 103
DB: 4 Gaps: 21

US-09-502-945-4 (1-2236) x US-09-562-737-24 (1-724)
QY 340 TCCAGGAAGCTCAGGAGGTGGCTGTGACCGCTCTGCACCCGCGGAGGCTCGGCTGAGT 399
||| : : : |||...||| |||
Db 155 AlaGluLysIleIleGluIleLysLeuTrpLys---GlyProLysGlyLeuGlyPheSer 173
||| : : : |||...||| |||
QY 400 GTGCGTGTGGC-----CTGGAGTTTGGCTGTGGCTCTTCACTCCCAAC 444
||| : : : |||...||| |||
Db 174 TyrAlaGlyGlyValGlyAsnGlnHisIleAlaGlyAspAsnSerIleTyrValThrLys 193
||| : : : |||...||| |||
QY 445 CTCATCAAGGGGTGAGCA---GACAGCGTGGGCTCAGGATGAGGAGCATGCTC 501
||| : : : |||...||| |||
```

```
Db 194 AspIleGluGlyGlyAlaAlaHisLysAspGluArgLeuGlnIleGlyAspLysIleLeu 213
||| : : : |||...||| |||
QY 502 CGGATCAATGGATATTCATCTCTCTCTGATCCCATGAGGAGGTGATCAACCTCATTCGA 561
||| : : : |||...||| |||
Db 214 PheValAsnSerValGlyLeuGluAspValGlyHisGluAspAlaValAlaAlaLeuLys 233
||| : : : |||...||| |||
QY 562 ---ACCAAGAAAACCTGTGCCATCAAGTG-----AGACACATCGCCTGATCCCGTG 612
||| : : : |||...||| |||
Db 234 HisThrAspValValTyrLeuLysValIleLysProSerAsnAlaTyrLeuSerAsp 253
||| : : : |||...||| |||
QY 613 AAAAGTCTCTCATGAGCCCTCACT-----TGCAGTATGTGGATCAGTTGTGTCTG 666
||| : : : |||...||| |||
Db 254 LysTyrAlaProProAspIleThrThrSerLeuSerGlnHisLeuAspAsnGluIleSer 273
||| : : : |||...||| |||
QY 666 ----- 666
Db 274 MetSerSerTyrLeuGlyThrAspTyrProAsnAlaMetThrProThrSerProArgArg 293
||| : : : |||...||| |||
QY 667 GAATCGGGGCGCTGCGAGCAGCCCTGGCTCCCTTGGAAATCGGAAAACACAGGAGAAG 726
||| : : : |||...||| |||
Db 294 GlnSerProValAlaLysAspLeuGly-----ArgGluAspIleProArg 309
||| : : : |||...||| |||
QY 727 AAGTCTTTCATCAGCCTGTGTA-----GGCTCCGAGGCTTGGCTGCAGCATTTCCAGC 780
||| : : : |||...||| |||
Db 310 GluProArgArgSerValIleHisArgGlySerThrGlyLeuThrPheAsnIleValGly 329
||| : : : |||...||| |||
QY 781 GGCCCCATCCAGAACGCTGCTTTATCAGCCATGTG-----AAACTGGCTCC 831
||| : : : |||...||| |||
Db 330 GlyGlu---AspGlyValGlyIlePheIleSerPheIleLeuAlaTyrGlyProAlaAsp 348
||| : : : |||...||| |||
QY 832 CTGCTCTCTGAGTGGGATGGAGATAGGGACACAGATCTCGAAGTCAATGCGCTGCAC 891
||| : : : |||...||| |||
Db 349 LeuSerGlyGlu-----LeuTyrLysGlyAspGlnIleLeuSerValAsnAlaValAsp 366
||| : : : |||...||| |||
QY 892 TTCTTAACCTGGATCACAGGAGGTGTAATGTGTGAAATAGCCAGCCGCTGACC 951
||| : : : |||...||| |||
Db 367 LeuArgAsnAlaSerHisGluAspAlaIleAlaLeuLysAsnAlaGlyGlu---Thr 385
||| : : : |||...||| |||
QY 952 ATCTCATTCTAGTCTGAGCTGCGGGAGCTGTTATGACACACGGGAGCGGCTGGCA 1011
||| : : : |||...||| |||
Db 386 ValThrIleIleAlaGln-----TyrLysPheGluGluTyrSerArgPhe 400
||| : : : |||...||| |||
QY 1012 GAGCGCGGCGAGCTGAGCTGAGCGGCGAGGCTT----- 1047
||| : : : |||...||| |||
Db 401 GluAlaLysGlyHisAspLeuArgGluGlnLeuMetAsnHisSerLeuGlySerGlyThr 420
||| : : : |||...||| |||
QY 1047 ----- 1047
Db 421 AlaSerLeuIleSerAsnProLysArgGlyPheTyrIleLysAlaLeuPheAspTyrAsp 440
||| : : : |||...||| |||
QY 1048 -----CTCATGCAGAGCGGCTGGGATGGAGTCCCAACAAGATCCTC 1089
||| : : : |||...||| |||
Db 441 LysThrLysLeuCysGlyPheLeuSerGlnAlaLeuSerMethHisPheGlyAspValLeu 460
||| : : : |||...||| |||
QY 1090 CAGGACGAGGAGGATGGAGCGGCAAGGAGAAAGAAATTCGCCAGAGGCGGAGAG 1149
||| : : : |||...||| |||
Db 461 HisValIleAsnAlaSerAspGluTyrTrpGln-----AlaGlnArgValHisSer 478
||| : : : |||...||| |||
QY 1150 GAAATGAGAGATACCGGAGGAGATGGAACAGATTGTAGAGGAGGAGGAAGATTAAAG 1209
||| : : : |||...||| |||
Db 479 AspSerGluThrAspArg-----IleGlyPheIleProSerLysArgArgSerGluArg 496
||| : : : |||...||| |||
QY 1210 AAGCAATGGGAA-----GAAGACTGGGGCTCA----- 1236
||| : : : |||...||| |||
Db 497 ArgGluTrpSerArgLeuLysThrLysAspTyrGlySerSerSerGlySerValGlyArg 516
||| : : : |||...||| |||
QY 1237 AAGCAACAGCTACTCTTGGCCCTAAACCATCACT---GCTAGGTACAC-----CCA 1284
||| : : : |||...||| |||
Db 517 GluAspSerValLeuSerTyrTrpThrValThrGlnMetGluValHisTyrTyrArgPro 536
||| : : : |||...||| |||
QY 1285 GTACCCCTTCGCAAGCAAAAGTATGATCAGGAGTGAACCTGAGCTCGAGCCCGCAGAT 1344
||| : : : |||...||| |||
Db 537 IleIleIleLeuGlyProThrAlaAspArgAlaAsnAspLeuLeuSerAspPhePro 556
||| : : : |||...||| |||
```



```

386 ProAlaLeuGlyPheProLysAlaLeuThrProAsnSerProProAlaMetValSerSer 405
QY 1354 GGAGGCACGAGGAGCAGGAGCAGCAGGATTCCGGAATATGAGGAAGCGCTTTGACCCC 1413
Db 406 SerSerProThrSerMetSerAlaTySerLeuSerSerLeuAsnMetGlyThrLeuPro 425
QY 1414 TACTCTATGTC-----ACCCAGAGCAGATCATGGGGAAGGATGTCGGCTCCACGC 1467
Db 426 ArgSerLeuTySerThrSerProArgGlyThrMetMetArg-----ArgArg 441
QY 1468 ATCAAGAAG-----GAGGGATCCTTAGACCTGGCC----- 1497
Db 442 LeuLysLysAspPheLysSerLeuSerLeuAlaSerSerThrValGlyLeuAla 461
QY 1497 ----- 1497
Db 462 GlyGlnValValHisThrGluThrThrGluValValLeuThrAlaAspProValThrGly 481
QY 1498 -----CTGGAAGCGCGTG-----GACTCCCCCATTTGGG 1527
Db 482 PheGlyIleGlnLeuGlnGlySerValPheAlaThrGluThrLeuSerSerPro----- 499
QY 1528 AAGGTGTCGTTCTCTGCTGATGAGCGGGAGCTGCTGAGCGCATGGTGGCATTGTG 1587
Db 500 ---ProIleSerTyIleGluAlaaspSerProAlaGluArgCysGlyValLeuGln 518
QY 1588 AAAGGGCAGCATATGCAATCAACGCGAAGATTGACAGACTACACCCCTGGCTGAG 1647
Db 519 IleGlyAspArgValMetAlaIleAsnGlyIleProThrGluAspSerThrPheGluGlu 538
QY 1648 GCTGACGCTCCCTGAGAAAGCGCTGGAATCAGGCGGGGACTCGATC----- 1695
Db 539 AlaAsnGlnLeuLeuArgaspSerSerIleThrSerLysValThrLeuGluIleGluPhe 558
QY 1696 GACCTGTGCTGCCCTCTGCCCC-----CCAAAGGAG 1728
Db 559 AspValAlaGluSerValIleProSerSerGlyThrPheHisValLysLeuProLysLys 578
QY 1729 TATGACGATGAGCTGACCTTCTTGCTGAAGTCC----- 1761
Db 579 HisSerValGluLeuGlyIleThrIleSerSerProSerSerArgLysProGlyAspPro 598
QY 1762 -----AAAAGGGGAACCAAAATTCACGCGTTAGGAACAGTGAGCTC 1803
Db 599 LeuValIleSerAspIleLysLysGlySerValAlaHisArgThrGlyThrLeuGluLeu 618
QY 1804 CGGCCCCACCTCGTG 1818
Db 619 GlyAspLysLeuLeu 623
RESULT 9
US-09-045-632-3
; Sequence 3, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: HUGANIR, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

Db	217	AlaGluAlaMetSerIleLeuLysGlnCysGlyGlnGluAlaThrLeuLeuIleClyr	236
QY	955	-----TCCATTGTAGCTGCAGCTGGC-----	975
Db	237	AspValSerAlaMetAspSerValAlaThrAlaSerGlyProLeuLeuValGluValAla	256
QY	976	-----CGGGAG	981
Db	257	LysThrProGlyAlaSerLeuGlyValAlaLeuThrThrSerValCysAsnLysGln	276
QY	982	CTGTTTCATGACAGCCGGAGCGCTGGCAGGAGCGCGCAGCGT-----	1026
Db	277	ValIleValIleAspLysIleLysSerAlaSerIleAlaAspArgCysGlyAlaLeuHis	296
QY	1027	-----GAGCTGCAGCGCAGGAGCTTCATCTCAGAGAGCGCTGCCGATCGAG	1074
Db	297	ValGlyAspHisIleLeuSerIleAspGlyThrSerMetGluTyrcysThrLeuAlaGlu	316
QY	1075	TCCAACAGATCCTC-----CAGGACGACGAGGATGGAGCGCAAGG	1119
Db	317	AlaThrGlnPheLeuGlyAsnThrThrAspGlnValLysLeuGluLeuProHisHis	336
QY	1120	AGAAAGAANAATCCCNAGAGGCGACAGAGGAANAATGAGATACCGGAAGGAGATGAA	1179
Db	337	GlnThrArgLeuAlaLeuLysGlyProAspHisValLysIleGlnArgSerAspArgGln	356
QY	1180	CAGATTCTAGAGGAGGAAGAAGTTTAAGAAGCAATGGGAAGAGACTGGGGCTCAAG	1239
Db	357	-----LeuProThr---AspProThrAlaSerSer	365
QY	1240	GAACAGCTACTCTTGCTAAACCACTACTGCTGAGGTACACCCAGTACCCCTTCGCAAG	1299
Db	366	GlnCysSerValHisThrAsnHisHisAsnProHisHisProAspHisCysArgVal	385
QY	1300	CCAAAG-----TATGATCAGGAGTGGAACTCTGAGTCGAGCCGCGAGACTCGAT	1353
Db	386	ProAlaLeuGlyPheProLysAlaLeuThrProAsnSerProProAlaMetValSerSer	405
QY	1354	GGAGCGCAGGAGGAGGAGCAGGAGATTTCGCGAAATATAGGAAGAGCTTTGACCCC	1413
Db	406	SerSerProThrSerMetSerAlaTyrcysLeuSerSerLeuAsnMetGlyThrLeuPro	425
QY	1414	TACTCTATGTC-----ACCCGACGACAGATCATGCGGAAGAGATGCCGGCTCTACGC	1467
Db	426	ArgSerLeuTyrcysThrThrProArgGlyThrMetMetArg-----ArgArg	441
QY	1468	ATCAAGAAG-----GAGGATCCTTAGACCTGGCC-----	1497
Db	442	LeuLysLysAspPheLysSerSerLeuSerLeuAlaSerSerThrValGlyLeuAla	461
QY	1497	-----	1497
Db	462	GlyGlnValValHisThrGluThrThrGluValValLeuThrAlaAspProValThrGly	481
QY	1498	-----CTGAAGCGGTGTG-----GACTCCCCCATTTGGG	1527
Db	482	PheGlyIleGlnLeuGlnGlySerValPheAlaThrGluThrLeuSerSerPro-----	499
QY	1528	AAGTGTGCTCTTCTGCTGTATGACGGGGAGCTGCTGAGCGGCATGTGGCATTTGTC	1587
Db	500	---ProLeuIleSerTyrcysGlyLeuAlaAspSerProAlaGluArgCysGlyValLeuGln	518
QY	1588	AAAGGGACGAGATCATGGCAATCAACGCGAAGATTGTGACAGACTACACCTGGCTGAG	1647
Db	519	IleGlyAspArgValMetalIleAsnGlyIleProThrGluAspSerThrPheGluGlu	538
QY	1648	GCTGACGCTCCCTCGCAGAGGCGCTGGAATCAGGCGGGGACTGGATC-----	1695
Db	539	AlaAsnGlnLeuLeuArgAspSerSerIleThrSerLysValThrLeuGluIleGluPhe	558
QY	1696	GACCTTGTGTTCCCTCTGCCCC-----CCAAAGGAG	1728

Db 559 AspValAlaGluSerValIleProSerSerGlyThrPheHisValLysLeuProLysLys 578
 QY 1729 TATGACGATGAGTCACCTTCCTTGCCTGAAGTCC----- 1761
 Db 579 HisSerValGluLeuGlyIleThrIleSerSerProSerSerArgLysProGlyAspPro 598
 QY 1762 -----AAAAGGGGAACCAAAATTCACGCCGTTAGGAACAGTGAGCTC 1803
 Db 599 LeuValIleSerAspIleLysLysGlySerValAlaHisArgThrGlyThrLeuGluLeu 618
 QY 1804 CGGGCCCACTCGTG 1818
 Db 619 GlyAspLysLeuLeu 623
 RESULT 10
 US-09-562-737-30
 ; Sequence 30, Application US/09562737
 ; Patent NO. 6428967
 ; GENERAL INFORMATION:
 ; APPLICANT: Herz, Joachim
 ; APPLICANT: Gotthardt, Michael
 ; TITLE OF INVENTION: LDL Receptor Signaling Pathways
 ; FILE REFERENCE: US/09/562,737
 ; CURRENT APPLICATION NUMBER: US/09/562,737
 ; CURRENT FILING DATE: 2000-05-01
 ; NUMBER OF SEQ ID NOS: 132
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 30
 ; LENGTH: 724
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: Sequence
 US-09-562-737-30
 Alignment Scores:
 Pred. No.: 1,766-08 Length: 724
 Score: 201.00 Matches: 113
 Percent Similarity: 39.96% Conservative: 100
 Best Local Similarity: 21.20% Mismatches: 218
 Query Match: 4.88% Indels: 102
 DB: 4 Gaps: 23
 US-09-502-945-4 (1-2236) x US-09-562-737-30 (1-724)
 QY 379 CCGAAGCCTCGCGCTGAGTGGCTGGTGGCTGGAGTT-----GGC 423
 Db 167 ProlysGlyTrpGlyPheSerIleAlaGlyGlyValGlyTyrGlnHisIleProGlyAsp 186
 QY 424 TGTGGCTCTTCACTCCCACTCATCAAGCGGTGAGGACAGACGCTCGGG---CTC 480
 Db 187 AsnSerIleAlaValThrLysIleIleGluGlyAlaAspHisLysAspGlyArgLeu 206
 QY 481 CAGGTAGGGACGAGATCGTCGGATCAATGATATTCATCTCTCTGTACCCATGAG 540
 Db 207 GlnIleGlyLysIleLeuAlaValAlaAsnSerValGlyPheGluAspValMetHisGlu 226
 QY 541 GAGGTCAATCAACTCATTCGA---ACCAAGAAAACGTGTCCATCAAGTGAGACACATC 597
 Db 227 AspAlaValGlyAlaLeuLysAsnThrTyrAspValValHisLeuLysValAlaLysPro 246
 QY 598 GGCCTGATCCCGTGAAAGCTCT-----CCTGATGAGCCCTCACTTGG---CAG 645
 Db 247 SerAsnAlaIleLeuSerAspSerTyrAlaProAspLysThrThrSerTyrSerGln 266
 QY 646 TATGTGATCAGTTGTGTGGAAATCTGGGGCGTGCAGGACGCTGGGC----- 696
 Db 267 HisLeuAspLeuGluIleSerHisSer-----SerTyrLeuGlyMetAspTyr 282
 QY 697 -----TCCCTCGAAATCGGAA----- 714
 Db 283 ProThrAlaMetThrProThrAsnProArgArgTyrSerProValAlaLysGlnLeuLeu 302

QY 715 -----AACAGGAGAGAGAGGCTTTCATCAGCCTGGTAGGCTCCCGA 756
Db 303 GlyGluGluAspIleProArgArgProArgIleValIleHisArg---GlyThrThr 321
QY 757 GGCCTTGCTGCAGCATTTCCAGCGGCCCATCCAGAACCTGGCTATTCATCAGCCAT 816
Db 322 GlyLeuGlyPheAsnIle---ValGlyValGluAspGlyGluGlyIlePheIleSerTrp 340
QY 817 GTGAACCTGGCTCCTCTGCTGCTGAGGTGGA---TTGGAGATAGGGGACCATGTC 873
Db 341 IleLeuAlaGlyGlyProAlaAspLeuTyrglyGluLeuArgLysGlyAspGlnIleAla 360
QY 874 GAAGTCAATGGCTCGACTTCTTAACCTGGATCAAGAGGCTGTAAATGCTGTA 933
Db 361 SerValasnGlyValasLeuArgasnaspSerHisGluGlnAlaAlaIleAlaLeuGlu 380
QY 934 AATAGCCGAGCCTGACCATCTCCATTGTAGCTGCGCGGAGCTGTTCATGACA 993
Db 381 AsnAla---GlyGlnThrValThrIlePheGlnTyrlsProGluGluTyrlsSerArg 399
QY 994 GACCGGAGCGCTGGCAGCGCGGCGAGCTGAGCTGCACGCGCAG----- 1041
Db 400 GlyGluAlaLysIleHisAspLeuArgGluHisLeuMetAsnSerSerLeuGlySerGly 419
QY 1042 -----GAGCTT 1047
Db 420 IleAlaSerLeuArgSerAsnProLysArgLysPheTyrlsArgAlaLeuPheAspTrp 439
QY 1048 CTCATCAGAGAGCGG-----CTGGCGATG-----GAGTCC 1077
Db 440 LeuLysThrLysaspGlyPheLeuSerMetAlaLeuSerPheHisPheGlyAspVal 459
QY 1078 ACAAGATCCTCCAGGAGCAGGAGATGGAGCGCGCAAGAGAGAAAGAAATTCGCCAG 1137
Db 460 AsnHisValIleAspAlaSerAspGluGluGlnTrpGlnAlaArgValHisSerAsp 479
QY 1138 AAGCGCAGCAGAGAA-----AATGAGATACCGGAAGAGATGCAACAGATT 1185
Db 480 ArgGluThrAspAspIleGlyPheIleProThrLysArgArgValGluArgGluTrp 499
QY 1186 GTAGAGGAGAGAGAGAGTTTAAGAGCAATGGAGAGAGCTGGGCTCAAGGAACAG 1245
Db 500 ValArgLeuLysAlaLysAspTrpGlySerTrpSerGlyserGlnGlyArgGluAspSer 519
QY 1246 CTACTCTTCCCTAAACCATCACT---GCTGAGGTACAC-----CCAGTACCCCTT 1293
Db 520 TyrLeuSerTyrgluThrValThrGlnMetAlaValHisTyrlsAlaArgProIleIleIle 539
QY 1294 CGCAAGCCAAAGTATGATCAGGAGTGGAACTGAGCTCGAGCCCGAGATGACCTGGAT 1353
Db 540 AspGlyProThrLysAspArgAlaAsnAspGluLeuLeuSerGluPheProAspLysPhe 559
QY 1354 GGAGGCGGAGGAGGAGGAGAGAGATTTCCGGAATATGAG---GAAGCTTTGAC 1410
Db 560 PheSerCysValProHisThrThrArgProGlyArgGluTyrgluIleAspGlyArgAsp 579
QY 1411 CCTACTCTATGTCACCCAGCAGATCATCGGGAAGATGCTCGGCTCTCTACGCATC 1470
Db 580 HisHisPheValSerArgGluLysMetIle---LysAspIleGlnAlaHisLysPhe 598
QY 1471 AACAAGGAGGATCCTTAGACCTGGCGCTGGAAGCGGTGTGAGCTCCCGCATTTGGAAG 1530
Db 599 IleLysAlaGlyGlnTyrlsAsnSerHisLeuTyrlsLeuThrSerValGlnSerValArgGlu 618
QY 1531 GTGGTC-----GTTTCTGCTGTGTATGAGCGG 1557
Db 619 ValMetGluGlnGlyLysHisCysIleLeuAspAsnSerAlaAsnAlaValArgArgLeu 638
QY 1558 GGAGCTGTGAGCGCGCATGTGTCATGTG-----AAAGGGGACGAGATC 1602
Db 639 GlnGlnAlaHisLeuHisProIleAlaIlePheArgArgProArgSerLeuGluAsnVal 658

QY 1603 ATGCAATCAACGGCAAGATTGTGACAGACTACACCTGGCTGAGGCTGAGCTGCC--- 1659
Db 659 LeuSerIleasnLysArgIleThrGluGluGlnThrArgLysAlaPheAspArgAlaThr 678
QY 1660 ---CTGCAGAGGCTCGAATCAGGCGGGGACTGGATC 1695
Db 679 LysValGluGlnGluPheThrGluCysPheSerTrpIle 691
RESULT 11
US-09-045-632-14
; Sequence 14, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: HUGANIR, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-045-632-14
Alignment Scores:
Pred. No.: 1,76e-08 Length: 604
Score: 200.50 Matches: 127
Percent Similarity: 35.89% Conservative: 93
Best Local Similarity: 20.72% Mismatches: 222
Query Match: 4.87% Indels: 171
DB: 3 Gaps: 25
US-09-502-945-4 (1-2236) x US-09-045-632-14 (1-604)
QY 364 CTGGACCTCTGCACCCCGAAGGC-----CTCGGCTGAGTGTGGTGGCTGGAG 417
Db 3 ValGluLeuMetLysLysGluGlyThrThrLeuGlyCysThrValSerGlyGlyLeasp 22
QY 418 TTGGCTGTGGCTCTTCTCATCTCCACCTCATCAAGCGGT---CAGGCGACAGAGCTC 474
Db 23 LysaspGlyLysProArgValSerAsnLeuArgGlnGlyIleAlaAlaSerAsp 42
QY 475 GGGCTCCAGGTAGGGGACGAGATGCTCGGATCAATGGATATTCATCTCCTCTGTACC 534

```
Db 43 GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg 62
QY 535 CATGAGGAGTCAACACCTCATTCGA---ACCAGAAACTGTCTCCATCAAGTGA 591
Db 63 HisAspGluIleIleSerLeuLysAsnValGlyGluArgValValLeuGluValGlu 82
QY 592 CACATCGGCGCTCATCCCGTGAAGAGTCTCTGTAGTGGCCCTCACTTGGCAGTATG 651
Db 83 Tyr-----GluLeuProValSerIleGlnGlySerSerValMetPheArgThrVal 100
QY 652 GATCATGTTCTCGAATCTGGGGCTGGCGAGGAGCGCTG-----GGTCCCTCGA 705
Db 101 GluValThrHisLysGluGlyAsnThrPheGlyPheValIleArgGlyAlaHis 120
QY 706 AATCGGAAACAAAGAGAGAGTCTTCATCAGCCTGGTAGGCTCCCGAGCGCTGCG 765
Db 121 AspAspArgAsnLysSerArgProValValIleThr----- 132
QY 766 TGCAGATTTCAGCGGCCCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAACCT 825
Db 133 CysValArgProGlyGlyProAspAspArgGluGlyThr-----IleLysPro 148
QY 826 GCCTCCCTCTGCTGCTGAGTGGATGGAGTAGGGACCATGTCGAATCAATGCG 885
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157
QY 886 GTCGACTTCTTAACCTGGATCACAGAGGCTGTAAATGCTGCTGAAA-----AAT 936
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177
QY 937 ACCCGAGCTGACCATC-----TCCATTGTAGCTGCAGCT 972
Db 178 GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer 197
QY 973 GCG----- 975
Db 198 GlyProLeuLeuValGluValAlaLysThrProGlyAlaSerLeuGlyValAlaLeuThr 217
QY 976 -----CGGAGCTGTTTCATGACAGACCGGAGCGGCTGGCAGAGCG 1017
Db 218 ThrSerValCysCysAsnLysGlnValIleValIleAspLysIleLysSerAlaSerIle 237
QY 1018 CGGACGCT-----GAGCTGCAGCGGAGGCTTCTC 1050
Db 238 AlaAspArgCysGlyAlaLeuHisValGlyAspHisIleLeuSerIleAspGlyThrSer 257
QY 1051 ATGCAGAGCGGCTGGCGATGGATCCACACAGATCTCTC-----CAGGAG 1095
Db 258 MetGluTyrCysThrLeuAlaGluAlaThrGlnPheLeuGlyAsnThrThrAspGlnVal 277
QY 1096 CAGCAGAGATGGAGCGGCAAGAGAGAAATGGCCAGAGGCGAGCAGAGAAAT 1155
Db 278 LysLeuGluIleLeuProHisGlnThrArgLeuAlaLeuLysGlyProAspHisVal 297
QY 1156 GAGAGATACCGAGAGGAGATGGAACAGATTCTAGAGGAGGAGAACTTTAAGAGCAA 1215
Db 298 LysIleGlnArgSerArgGln-----LeuPro 307
QY 1216 TGGNAGAGACTGGGGCTCAAGAACAGACTACTCTTCTTAAACCATCACTGCTGAG 1275
Db 308 Trp---AspProTrpAlaSerSerGlnCysSerValHisThrAsnHisHisAsnPro 326
QY 1276 GTACACCCAGTACCCCTTCGCAAGCCAAAG-----TATGATCAGGAGTGAACCTGAG 1329
Db 327 HisHisProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsn 346
QY 1330 CTCGAGCCCGCAGATGACCTGGATGGAGCGGAGGAGGAGGAGGAGGATTCGCG 1389
Db 347 SerProAlaMetValSerSerSerSerProThrSerMetSerAlaTyrSerLeuSer 366
QY 1390 AAATATGAGGAGGCTTTGACCCCTACTCTATGTTCT-----ACCCAGCAGCATCATG 1443
Db 367 SerLeuAsnMetGlyThrLeuProArgSerLeuTyrSerThrSerProArgGlyThrMet 386
```

```
QY 1444 GGAAGGATGTCCGCTCCTACGATCAAGAAG-----GAGGGATCCTTTAGACCTG 1494
Db 387 MetArg-----ArgArgLeuLysLysAspPheLysSerSerLeuSerLeu 402
QY 1495 GCC----- 1497
Db 403 AlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrGluValVal 422
QY 1498 -----CTGGAAGCGGTGTG----- 1512
Db 423 LeuThrAlaAspProValThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThr 442
QY 1513 -----GACTCCCCCATGGGAAGGTGCTTCTGCTGTGTATGACGGGGAGCT 1563
Db 443 GluThrLeuSerSerPro-----ProLeuIleSerTyrIleGluAlaAspSerPro 459
QY 1564 GCTGAGCGGCATGTGGCATTGTGAAGGGGACGAGATCATGGCAATCAACGGCAAGATT 1623
Db 460 AlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIlePro 479
QY 1624 GTGACAGACTACACCTGGCTGAGCTGACGCTGCCCTGCAGAGGCTTGAATCAGGGC 1683
Db 480 ThrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerSerIleThrSer 499
QY 1684 GGGGACTGGATC-----GACCTTGTGTTGCCCTCTGCCCTC----- 1719
Db 500 LysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerSerGlyThr 519
QY 1720 -----CCAAAGGAGTATGACGATGACGCTGACCTTCTTGTGTAAGTCC--- 1761
Db 520 PheHisValLysLeuProLysLysHisSerValGluLeuGlyIleThrIleSerSerPro 539
QY 1762 -----AAAAGGGGAAACCAAAAT 1779
Db 540 SerSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlySerValAla 559
QY 1780 CACCGGTAGAAACAGTACGCTCGGCGCCACCTCGTG 1818
Db 560 HisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572
RESULT 12
US-09-045-632-15
; Sequence 15, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Huginir, Richard L.
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0; Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
```


REFERENCE/DOCKET NUMBER: 48147/1699-CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 702 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-045-632-15

Alignment Scores:

Pred. No.: 1,91e-08 Length: 702

Score: 200.50 Matches: 127

Percent Similarity: 35.89% Conservative: 93

Best Local Similarity: 20.72% Mismatches: 222

Query Match: 4.87% Indels: 171

DB: 3 Gaps: 25

US-09-502-945-4 (1-2236) x US-09-045-632-15 (1-702)

```

QY 364 CTGGACGCTGCACCCGGAAGC-----CTGGCCGTGAGTGGCGGTGGCTGGAG 417
Db 3 ValGluLeuMetLysLysGluGlyThrThrLeuGlyCysThrValSerGlyGlyIleAsp 22
QY 418 TTGGCTGTGGCTCTTCATCTCCACCTCATCAAGCGGT--CAGCGACAGCGTC 474
Db 23 LysAspGlyLysProArgValSerAsnLeuArgGlnGlyGlyIleAlaAlaArgSerAsp 42
QY 475 GGGCTCCAGTGGGAGGAGATCGTCGGATCAATGGATATTCGATCTCCCTCCGTACC 534
Db 43 GlnLeuAspValGlyAspPyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg 62
QY 535 CATGAGGAGGTATCAACCTCATTCGA---ACCAGAAACTGTGTCCATCAAGTGAGA 591
Db 63 HisAspGluIleSerLeuLeuLysAsnValGlyGluArgValValLeuGluValGlu 82
QY 592 CACATCGGCTGATCCCGTGAAGAGCTCTCTGTATGAGCCCTCACTTGGCAGTATGG 651
Db 83 Tyr-----GluLeuProValSerIleGlnGlySerValMetPheArgThrVal 100
QY 652 GATCAGTTTGTGCGAARTCTGGGCGTGGCGAGGAGCGCTG-----GGCTCCCTGGA 705
Db 101 GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyAlaHis 120
QY 706 AATCGGGAACAAAGAGAGAGAGTCTTCATCAGCTCGTGGTCCCGAGCGCTGGC 765
Db 121 AspAspArgAsnLysSerArgProValValIleThr----- 132
QY 766 TGACGATTTCCAGCGGCCCCATCCAGAACCTTGGCATCTTTATCAGCCATGTGAACCT 825
Db 133 CysValArgProGlyLysProAspArgGluGlyThr-----IleLysPro 148
QY 826 GGCTCCCTGCTGCTGAGGTGGGATGGAGATAGGGGACCATGATTCGAAGTCAATGCG 885
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157
QY 886 GTCGCTTCTTAACCTGATCAACAGGAGGCTGTAATGTGCTGAAA-----AAT 936
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177
QY 937 AGCCGAGCGCTGACCATC-----TCCATTGTAGTCGAGCT 972
Db 178 GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer 197
QY 973 GGC----- 975
Db 198 GlyProLeuValGluValAlaLysThrProGlyAlaSerLeuGlyValAlaLeuThr 217
QY 976 -----CGGAGCTGTTCATGACAGACCGGGAGCGGTGGCAGAGCGG 1017
          ::::::::::: ::::: ::::: :::::

```

```

Db 218 ThrSerValCysCysAsnLysGlnValIleValIleAspLysLysIleLysSerAlaSerIle 237
QY 1018 CGGCAGCGT-----GAGCTGCGAGCGCGAGGAGCTTCTC 1050
Db 238 AlaAspArgCysGlyAlaLeuHisValGlyAspHisIleLeuSerIleAspGlyThrSer 257
QY 1051 ATGCAGAAAGCGGTGGCGATGGAGTCCCAACAAGATCCTC-----CAGGAG 1095
Db 258 MetGluTyrCysThrLeuAlaGluAlaThrGlnPheLeuGlyAsnThrThrAspGlnVal 277
QY 1096 CAGCAGGAGATGGAGCGGCAAGAGAAAGAAATTCCTCCAGAGGAGGAGGAGGAGAAAT 1155
Db 278 LysLeuGluIleLeuProHisHisGlnThrArgLeuAlaLeuLysGlyProAspHisVal 297
QY 1156 GAGAGATACCGGAGGAGATGGAAACAGATTCTAGAGGAGGAAGAGATTAAAGAGCAA 1215
Db 298 LysIleGlnArgSerAspArgGln-----LeuPro 307
QY 1216 TGGGAAGAAGACTGGGCTCAAGGAACAGACTACTTGGCTAAACCATCATCTGCTGAG 1275
Db 308 Trp---AspProTrpAlaSerSerGlnCysSerValHisThrAsnHisHisAsnPro 326
QY 1276 CTACACCCAGTACCCCTTCGCAAGCCAAAG-----TATGATCAGGGAGTGGAACTCAG 1329
Db 327 HisHisProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsn 346
QY 1330 CTCGAGCCGCGAGATGACCTGGATGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGTTCGG 1389
Db 347 SerProAlaMetValSerSerSerSerProThrSerMetSerAlaTyrSerLeuSer 366
QY 1390 AAATATAGGAAAGCTTTGACCCCTACTCTATGTC-----ACCCAGAGCAGATCATG 1443
Db 367 SerLeuAsnMetGlyThrLeuProArgSerLeuTyrSerThrSerProArgGlyThrMet 386
QY 1444 GGAAGGATGTCGGCTCCTACGATCAAGAG-----GAGGGATCCTTAGACCTG 1494
Db 387 MetArg-----ArgArgLeuLysLysLysAspPheLysSerLeuSerLeu 402
QY 1495 GCC----- 1497
Db 403 AlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrThrGluValVal 422
QY 1498 -----CTGGAAGCGGTGTG----- 1512
Db 423 LeuThrAlaAspProValThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThr 442
QY 1513 -----GACTCCCGCATTTGGGAAGGTGCTGCTTCTGCTGTATGAGCGGGAGCT 1563
Db 443 GluThrLeuSerSerPro-----ProLeuIleSerTyrIleGluAlaAspSerPro 459
QY 1564 GCTGAGCGCGCATGTGGCATTTGTAAAGGGGACGAGATCATGGCAATCAACGGCAAGATT 1623
Db 460 AlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIlePro 479
QY 1624 GTGACAGACTACACCTGCTGGCTGAGCTGACCTGCCCTGCGAGAGGCGCTGGAATCAGGGC 1683
Db 480 ThrGluAspSerThrPheGluAlaAsnGlnLeuLeuArgAspSerSerIleThrSer 499
QY 1684 GGGAGCTGGATC-----GACCTGTGTTGCTGCTGCCGCC----- 1719
Db 500 LysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerSerGlyThr 519
QY 1720 -----CCAAAGGAGTATGACGATGAGCTGAGCTTCTTCTGCTGAAGTCC--- 1761
Db 520 PheHisValLysLeuProLysLysHisSerValGluLeuGlyIleThrIleSerSerPro 539
QY 1762 -----AAAGGGGAACCAAAAT 1779
Db 540 SerSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlySerValAla 559
QY 1780 CACGGTTAGGAACAGTACGCTCGGCCCCCACCCTCGTG 1818
Db 560 HisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572

```

RESULT 13

US-09-045-632-16
Sequence 16, Application US/09045632
Patent No. 6001575
GENERAL INFORMATION:
APPLICANT: Huanir, Richard L.
APPLICANT: Dong, Hualing
TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
TITLE OF INVENTION: GRIP-RELATED MOLECULES
NUMBER OF SEQUENCES: 105
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/045,632
FILING DATE: 19-MAR-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/041,016
FILING DATE: 19-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Corless, Peter F.
REGISTRATION NUMBER: 33,860
REFERENCE/DOCKET NUMBER: 48147/1699-CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-045-632-16

Alignment Scores:

Pred. No.:	2,34e-08	Length:	1018
Score:	200.50	Matches:	127
Percent Similarity:	35.89%	Conservative:	93
Best Local Similarity:	20.72%	Mismatches:	222
Query Match:	4.87%	Indels:	171
DB:	3	Gaps:	25

US-09-502-945-4 (1-2236) x US-09-045-632-16 (1-1018)

QY	364	CTGAGCGCTGACCCCGAAGGC-----CTGGCGCTGAGTGGCGTGGTGGCGCTGGAG	417
Db	3	ValGluLeuMetLysLysGluGlyThrThrLysGlyCysThrValSerGlyGlyLeuA	22
QY	418	TTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGCGGT---CAGCGACACGCGTC	474
Db	23	LysAspGlyLysProArgValSerAsnLeuArgGlnGlyGlyLeuAlaAlaArgSerAsp	42
QY	475	GGGCTCCAGTAGGGAGGAGATCGTCGGGATCAATGGATATTCCTCTCTCTGCTACC	534
Db	43	GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg	62
QY	535	CATGAGGAGGTATCAACCTCATTCGA---ACCAAGAAACTGTGTCATCAAGTGAGA	591
Db	63	HisAspGluIleIleSerLeuLeuLysAsnValGlyGluArgValValLeuGluValGlu	82
QY	592	CACATCGGCGCTGATCCCGTGAAGAGCTCTCTGTATGAGCGCCCTCACTTGGCGAGTATGTG	651

Db	83	Tyr-----GluLeuProValSerIleGlnGlySerValMetPheArgThrVal	100
QY	652	GATCAGTTTGTGTCGGAATCTGGGGCGTGCAGGAGCAGCTG-----GGCTCCCTGGA	705
Db	101	GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyGlyAlaHis	120
QY	706	AATCGGAAACAAAGAGAGAAAGGTCTTCATCAGCTCGGTAGGCTCCCGAGCGCTTGGC	765
Db	121	AspAspArgAsnLysSerArgProValValIleThr-----	132
QY	766	TGCAGCATTTCCAGCGCCCATCCAGAACCTCGGATCTTTATCAGCCATGTGAACCT	825
Db	133	CysValArgProGlyGlyProAspArgGluGlyThr-----IleLysPro	148
QY	826	GGCTCCCTGTCTGCTGAGGTGGATTTGGAGATAGGGGACGACAGATTCTCGAAGTCAATGGC	885
Db	149	-----GlyAspArgLeuLeuSerValaspGly	157
QY	886	GTGACTTCTTAACCTGGATCAAGAGGCGTGTAAATGTGCTGAAA-----AAT	936
Db	158	IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln	177
QY	937	AGCCGAGCGCTGACCATC-----TCCATTGTAGCTGCGAGCT	972
Db	178	GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer	197
QY	973	GGC-----	975
Db	198	GlyProLeuLeuValGluValAlaLysThrProGlyAlaSerLeuGlyValAlaLeuThr	217
QY	976	-----CGGAGCTGTTTCATGCACAGACCGGAGCGGCTGGCAGAGCGG	1017
Db	218	ThrSerValCysCysAsnLysGlnValIleValIleAspLysIleLysSerAlaSerIle	237
QY	1018	CGCGACGGT-----GAGCTGCAGCGGAGGAGCTTCTC	1050
Db	238	AlaAspArgCysGlyAlaLeuHisValGlyAspHisIleLeuSerIleAspGlyThrSer	257
QY	1051	ATGCAGAACGGCTGGCGATGGATCCCAACAGATCCTC-----CAGGAG	1095
Db	258	MetGluTyrCysThrLeuAlaGluAlaThrGlnPheLeuGlyAsnThrThrAspGlnVal	277
QY	1096	CAGCAGGATGGCGGCAAGAGAGAAATTTGCCAGAGGACGACGAGAGAAAT	1155
Db	278	LysLeuGluIleLeuProHisHisGlnThrArgLeuAlaLeuLysGlyProAspHisVal	297
QY	1156	GAGAGATACCGNAGGAGATGGACAGATCTAGAGGAGAGAGAGATTAAAGAGCAA	1215
Db	298	LysIleGlnArgSerAspArgGln-----LeuPro	307
QY	1216	TGGGAAGAAGACTGGGGCTCAAAGGAACAGTACTCTTCCCTAAACCATCATCTGCTGAG	1275
Db	308	Trp---AspProTrpAlaSerSerGlnCysSerValHisThrAsnHisHisAsnPro	326
QY	1276	GTACACCCAGTACCCCTTCGCAAGCCAAAG-----TATGATCAGGGAGTGAACCTGAG	1329
Db	327	HisHisProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsn	346
QY	1330	CTGAGCGCGGAGATGACCTGGATGGAGCGGAGGAGGAGGAGGAGGAGGATTTCGGG	1389
Db	347	SerProProAlaMetValSerSerSerSerProThrSerMetSerAlaTyrSerLeuSer	366
QY	1390	AAATATGAGGAAGCTTTTGACCCCTACTCTATGTTC-----ACCCAGAGCAGATCATG	1443
Db	367	SerLeuAsnMetGlyThrLeuProArgSerLeuTyrSerThrSerProArgGlyThrMet	386
QY	1444	GGGAAGGATGTCGGGCTCTACGATCATCAAGAA-----GAGGGATCTCTTAGACCTG	1494
Db	387	MetArg-----ArgArgLeuLysLysLysAspPheLysSerSerLeuSerLeu	402
QY	1495	GCC-----	1497

```
Db 403 AlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrThrGluValVal 422
QY 1498 -----CTGGAAGCGGTGTG-----1512
Db 423 LeuThrAlaAspProValThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThr 442
QY 1513 -----GACTCCCGCATTTGCTGCTGTGTATGAGCGGGAGCT 1563
Db 443 GluThrLeuSerPro-----ProLeuIleSerThrIleGluAlaAspSerPro 459
QY 1564 GCTGAGCGGATGTCATGTGAAGGAGGAGATGTCGCAATCAACGGCAAGATT 1623
Db 460 AlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIlePro 479
QY 1624 GTGACAGACTACCCCTGGCTGAGCTGAGCTGCCCTGCGAGAAGGCTGGAATCAGGCG 1683
Db 480 ThrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerIleThrSer 499
QY 1684 GGGGACTGGATC-----GACCTTGTGTTGCCCTGCTCCCGC-----1719
Db 500 LysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerSerGlyThr 519
QY 1720 -----CCAAAGGAGTATGACGATGAGCTGACCTTCTTGTGCAAGTCC---1761
Db 520 PheHisValLysLeuProLysLysHisSerValGluLeuGlyIleThrIleSerSerPro 539
QY 1762 -----AAAGGGGAAACCAAAATT 1779
Db 540 SerSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlySerValAla 559
QY 1780 CACGCGTTAGGAACAGTAGCTAGCTCGGCGCCACCTCGTG 1818
Db 560 HisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572
```

RESULT 14

```
US-09-045-632-32
; Sequence 32, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Hugenir, Richard L.
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045.632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1061 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-045-632-32

Alignment Scores:
Pred. No.: 2,39e-08 Length: 1061
Score: 200.50 Matches: 127
Percent Similarity: 35.89% Conservative: 93
Best Local Similarity: 20.72% Mismatches: 222
Query Match: 4.87% Indels: 171
DB: 3 Gaps: 25

US-09-502-945-4 (1-2236) x US-09-045-632-32 (1-1061)
QY 364 CTGACCGCTGACACCCGAGGC-----CTGGCGCTGAGTGTGCTGGTGCCTGGAG 417
Db 3 ValGluLeuMetLysLysGluGlyThrThrLeuGlyCysThrValSerGlyGlyIleAsp 22
QY 418 TTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGGCGGT---CAGCGACAGACGCTC 474
Db 23 LysAspGlyLysProArgValSerAsnLeuArgGlnGlyIleAlaAlaArgSerAsp 42
QY 475 GGGCTCAGTAGGAGGAGAGATCGCCGGATCAATGGATATTCATCTCTCTCTGTACC 534
Db 43 GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg 62
QY 535 CATGAGGAGGTCTACACCTCATTCGA--ACCAAGAAACTGTGCTCATCAAGTGAGA 591
Db 63 HisAspGluIleIleSerLeuLeuLysAsnValGlyGluArgValValLeuGluValGlu 82
QY 592 CACATCGCCCTGATCCCGTGAAGGCTCTCTGATGAGCCCTCACTTGGCAGTATGTG 651
Db 83 Tyr-----GluLeuProProValSerIleGlnGlySerSerValMetPheArgThrVal 100
QY 652 GATCAGTTTGTGCGGAATCTGGGGCGTGCAGAGGAGCGCTG-----GGCTCCCTGGA 705
Db 101 GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyAlaHis 120
QY 706 AATCGGGAAACAGAGAGAGGCTCTTCATCAGCCCTGCTAGGCTCCCGAGGCTTGGC 765
Db 121 AspArgAsnLysSerArgProValIleThr-----132
QY 766 TGCAGCATTTCCAGCGGCCCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAACCT 825
Db 133 CysValArgProGlyGlyProAspAspArgGluGlyThr-----IleLysPro 148
QY 826 GGCTCCCTGTCTGCTGAGTGGGATGAGATAGGGGACCAGATTGTCGAGTCAATGGC 885
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157
QY 886 GTCGACTTCTTAACCTGGATCACAGGAGGTGTAATGTGCTGAAA-----AAT 936
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177
QY 937 AGCGCGACCTGACCATC-----TCCATTGTAGCTGCAGCT 972
Db 178 GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer 197
QY 973 GGC-----975
Db 198 GlyProLeuLeuValGluValAlaLysThrProGlyAlaSerLeuGlyValAlaLeuThr 217
QY 976 -----CGGAGCTGTTATGACACCGGAGCGGCTGGAGAGCGG 1017
Db 218 ThrSerValCysCysAsnLysGlnValIleValIleAspLysIleLysSerAlaSerIle 237
QY 1018 CGGCAGCGT-----CAGCTGCGAGCGGAGGCTTCTC 1050
Db 238 AlaAspArgCysGlyAlaLeuHisValGlyAspHisIleLeuSerIleAspGlyThrSer 257
QY 1051 ATGCAGAGCGGCTGGCGATGGAGTCCAAACAAGATCCTC-----CAGGAG 1095
```

```
Db 258 MetGluTrpCysThrLeuAlaGluAlaThrGlnPheLeuGlyAsnThrThrAspGlnVal 277
QY 1096 CAGCAGGAGATGGCGCGCAAGAGGAGAAAGAAATGGCCAGAGGCGAGAGGAAAT 1155
Db 278 LysLeuGluIleLeuProHisGlnThrArgLeuAlaLeuLysGlyProAspHisVal 297
QY 1156 GAGAGATACCGGAGGAGATGGAAGATGTTAGAGGAGGAGAGAGAGATTTAAGAGCAA 1215
Db 298 LysIleGlnArgSerAspArgGln-----LeuPro 307
QY 1216 TGGGAAGAAGACTGGGCTCAAGAACAGCAGCTACTCTTCCCTAAACCATCACTGCTGAG 1275
Db 308 Trp---AspProTrpAlaSerSerGlnCysSerValHisThrAsnHisHisAsnPro 326
QY 1276 GTACACCCAGTACCCTTCGAGCCAAAG-----TATGATCAGGGAGTGAACCTGAG 1329
Db 327 HisHisProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsn 346
QY 1330 CTCGAGCCCGCAGATGACCTGGATGGAGCCAGGAGGAGGAGGAGGAGGATTTCCGG 1389
Db 347 SerProAlaMetValSerSerSerSerProThrSerMetSerAlaTrpSerLeuSer 366
QY 1390 AAATATGAGGAGGCTTACCCCTACTCTATGTTCT-----ACCCAGAGCAGATCATG 1443
Db 367 SerLeuAsnMetGlyThrLeuProArgSerLeuTrpSerThrSerProArgGlyThrMet 386
QY 1444 GGGAGGATGTCGGCTCTCAGCATCAAGAAG-----GAGGGATCCTTACAGCTG 1494
Db 387 MetArg-----ArgArgLeuLysLysLysAspPheLysSerSerLeuSerLeu 402
QY 1495 GCC-----1497
Db 403 AlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrThrGluValVal 422
QY 1498 -----CTGGAAGCGGTGTG-----1512
Db 423 LeuThrAlaAspProValThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThr 442
QY 1513 -----GACTCCCATGGGAAGGTGCTGTTCTGCTGTGTATGAGCGGGGAGCT 1563
Db 443 GluThrLeuSerSerPro-----ProLeuIleSerTrpIleGluAlaAspSerPro 459
QY 1564 GCTGAGCGGATGTGGCTATTGAAAGGGAGCAGATCATGGCAATCAACCGCAAGATT 1623
Db 460 AlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIlePro 479
QY 1624 GTGACAGACTACCCCTGGCTGAGCTGACGCTGCCCTGCAGAAAGGCTGGAATCAGGGC 1683
Db 480 ThrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerSerIleThrSer 499
QY 1684 GGGACTGGATC-----GACCTTGTGGTGGCTGCTGCC-----1719
Db 500 LysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerSerGlyThr 519
QY 1720 -----CCAAAGGAGTATCAGCATGAGCTGACCTTCTTCTGCTGAAGTCC---1761
Db 520 PheHisValLysLeuProLysHisSerValGluLeuGlyIleThrIleSerSerPro 539
QY 1762 -----AAAAGGGGAAACCAATTT 1779
Db 540 SerSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlySerValAla 559
QY 1780 CACCGGTTAGAAACAGTACGCTCGGCGCCCACTCGTG 1818
Db 560 HisArgThrGlyThrLeuGluLeuGlyAspLysLysLeu 572
```

RESULT 15

US-09-562-737-22

; Sequence 22, Application US/09562737

; Patent No. 6428967

; GENERAL INFORMATION:

; APPLICANT: Herz, Joschim

```
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: OTS0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-22
```

```
Alignment Scores: 2,36e-08 Length: 724
Pred. No.: 199,50 Matches: 122
Score: 34.08% Conservative: 76
Best Similarity: 21.00% Mismatches: 149
Best Local Similarity: 4.84% Indels: 235
Query Match: 4 Gaps: 24
DB:
US-09-502-945-4 (1-2236) x US-09-562-737-22 (1-724)
QY 247 AATGAACCCAGCGCTGCTCTCTGTTTGTATGCCATTGGCGCTGATC-----294
Db 24 HisSerProAlaHisLeuProAsnPheAlaAsnSerProProValIleValAsnGlyAsp 43
QY 295 -----CCACTGAAGCACCAGGTGAA-----TAT 318
Db 44 ThrLeuGluAlaProGlyTrpGluHisGlnValAsnGlyThrGluGlyGluMetIleTrp 63
QY 319 GATCAGCTGACCCCGCGCTCCAGGAAGCTGAAGGAGGTGCTGGACCGCTCTGCAC 378
Db 64 GluGluIleThrLeuGluArgGlyLys-----72
QY 379 CCGGAAGCGCTCGCGCTGAGTGTGCGTGT-----GGCTTGGAGTTTGGC 423
Db 73 ---SerGlyLeuGlyPheSerIleAlaGlyLeuThrAspAsnProHisIleGlyAspAsp 91
QY 424 TGTGGGCTCTTCTCCACCTCATCAAGGCGGTGATGATTTCCATCTCCTCTATCCCATGAG 540
Db 92 MetSerIlePheIleThrLysIleIleProAsnGlyAlaAlaGlnAspGlyArgLeu 111
QY 481 CAGTAGGGGACGACGATCGTCGGATCATATGATATTTCCATCTCCTCTATCCCATGAG 540
Db 112 GlnValAsnAspSerIleLeuPheValAsnArgValAspValArgGluValThrHisSer 131
QY 541 GAGTCTATCAACCTCATTCGACCAACGAAACACTGTGTCCATCAAGTGACACATCGGC 600
Db 132 SerAlaValGluAlaLeuLysGluAlaGlyThrIle-----ValArg 145
QY 601 CTGATCCCGCTGAAAGCTCTCTCTGATGAGCCCTCATCTTGGCAGTATGGGATCAGTTT 660
Db 146 LeuTrpValMetArgArgValProPro-----154
QY 661 GTGTGGGAATCTGGGGCGTGGGAGGCGCTGGGCTCCCTCGGAAATCGGAAACAAG 720
Db 155 -----Ala 155
QY 721 GAGAAGAGGTCTTCATCAGCCTGGTA---GGCTCCGGAGGCGCTTGGCTGCACATTTCC 777
Db 156 GlutylsIleIleGluIleThrLeuIleLysGlyProLysGlyLeuGlyTrpSerIleAla 175
QY 778 AGCGGCCCTCCAGAG-----CTGGC-----ATCTTATCACCCTATGTGAA 822
Db 176 GlyGlyValGlyAsnGlnAlaIleProGlyAspAsnSerIleTrpValAspLysIleIle 195
QY 823 CCTGGCTCCCTGCTGCTGAGGTGGGA---TTGGAGATAGGGACCAAGATTGTCGAAGTC 879
Db 196 GluGlyGlyAlaAlaHisGluAspGlyArgLeuGlnIleGlyAspLysPheLeuAlaVal 215
```



```
Alignment Scores:
Pred. No.: 2,63e-93 Length: 268
Score: 1350.00 Matches: 262
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 32.77% Indels: 0
DB: 10 Gaps: 0

US-09-502-945-4 (1-2236) x US-09-739-907-114 (1-268)
Qy 964 GCTGAGCTGCGCGGAGCTGTTATGACAGCCGGGAGCGCTGGCAGAGCGCGGCAG 1023
Db 6 AlaAlaAlaGlyArgGluLeuPheMetThrAspArgGluArgLeuAlaGluAlaArgGln 25
Qy 1024 CGTGAGCTGCAGCGCAGAGAGCTTCTCATGAGAACCGCGCTGGCGATGGATCCACAG 1083
Db 26 ArgGluLeuGlnArgGlnGluLeuMetGlnLysArgGluLeuAlaMetGluSerAsnLys 45
Qy 1084 ATCCCTCCAGGACGACGAGATGAGCGGCAAGGAGAAAGAAATTCGCCAGAGGCA 1143
Db 46 lleLeuGlnGlnGlnGlnGlnMetGluArgGlnArgLysGluilleAlaGlnLysAla 65
Qy 1144 GCAGAGGAAATAGAGATACCGGAGGAGATGGAACAGATGTAGAGGAGGAAGAGAG 1203
Db 66 AlaGluGluAsnGluArgTyArgLysGluMetGluGlnIleValGluGluGluLys 85
Qy 1204 TTAAAGAACCAATGGGAAGAGACTGGGGCTCAAGAGGACACTACTCTTGCCTAAACC 1263
Db 86 PheLysLysGlnTrpGluAspTrpGlySerLysGluGlnLeuLeuProLysThr 105
Qy 1264 ATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCACAAGATGATGATCAGGAGTGAA 1323
Db 106 IleThrAlaGluValHisProValProLeuArgLysProLysTyArgGlnGlyValGlu 125
Qy 1324 CTGAGCTGAGCCCGCAGATGACCTGGATGAGGACGAGGACGAGGAGAGCAGGAT 1383
Db 126 ProGluLeuGluProAlaAspLeuAspGlyGlyThrGluGluGlnGlyGluGlnAsp 145
Qy 1384 TTCCGGAATATAGGAAGGCTTTGACCCCTACTCTATGTTACCCCGCAGAGCATG 1443
Db 146 PheArgLysTyArgGluGlyPheAspProTySerMetPheThrProGluGlnIleMet 165
Qy 1444 GGAAGGATGTCGGGCTCCTACGCATCAAGAAGGAGGATCCTTAGACTGGCCCTGGAA 1503
Db 166 GlyLysAspValArgLeuLeuArgLysLysGlySerLeuAspLeuAlaLeuGlu 185
Qy 1504 GCGGTGTGAGTCCCGCATTCGGAAGGTGCGTTCGTGCTGTATGAGCGGGAGCT 1563
Db 186 GlyGlyValAspSerProIleGlyLysValValSerAlaValTyGluArgGlyAla 205
Qy 1564 GCTGAGCGCATGGTGGCATTCGAAAGGAGGACGATCATGGCAATCAACGGCAAGATT 1623
Db 206 AlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLysIle 225
Qy 1624 GTGACAGCTACACCTGCTGAGGCTGACGCTGCCCTGCAGAAGCCCTGGAATCAGGCG 1683
Db 226 ValThrAspTyThrLeuAlaGluAlaAspAlaAlaLeuGlnLysAlaTrpAsnGlnGly 245
Qy 1684 GGGGACTGATGCACTTGTGTTGCTGCTGCCCTGCCCGCCAAAGAGTATGACCATGCTG 1743
Db 246 GlyAspTrpIleAspLeuValValAlaValCysProProLysGluTyAspAspGluLeu 265
Qy 1744 ACCTTC 1749
Db 266 ThrPhe 267

RESULT 2
US-09-925-299-778
; Sequence 778, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

US-09-925-299-778
; Sequence 778, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 778
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-502-945-4 (1-2236) x US-09-925-299-778 (1-120)
Qy 1315 GGAGTGAACCTGAGCTCGAGCCGCGAGATGACCTGGATGGAGCAGGAGGACGGA 1374
Db 23 GlyLeuArgProValLeuGlnPro-----ArgGlnGly 33
Qy 1375 GAGCAGGATTCGCGAATATGAGGAAGGCTTGACCCCTACTCTATGTTACCCCGAG 1434
Db 34 ---GlnAspPheArgLysTyArgLysGluGluGlyPheAspProTySerMetPheThrProGlu 52
Qy 1435 CAGATCATGGGAAGGATGTCGGCTCTAGCATCAAGAGGAGGATCCTTAGACCTG 1494
Db 53 GlnIleMetGlyLysAspValArgLeuLeuArgLysGluGlySerLeuAspLeu 72
Qy 1495 GCCCTGGAAGCGGTGGACTCCCGCATTCGCCATGGGAAGGTGCTGCTGTGTATGAG 1554
Db 73 AlaLeuGluGlyGlyValAspSer**IleGlyLysValValSerAlaValTyGlu 92
Qy 1555 CGGGAGCTGCTGAGCGGCATGTCGATGTGAAGGGGACGAGATCGGCAATCAAC 1614
Db 93 ArgGlyAlaAlaGluArgHisGlyIleValLysGlyAspGluIleMetAlaIleAsn 112
Qy 1615 GGCAAGATTGTGACAGACTACACC 1638
Db 113 GlyLysIleValThrAspTyThr 120

RESULT 3
US-09-925-299-778
; Sequence 778, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 778
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```

;
; NAME/KEY: SITE
;
; LOCATION: (81)
;
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-778

```

Alignment Scores:	
Pred. No.:	1,268-25
Score:	47.00
Percent Similarity:	85.04%
Best Local Similarity:	85.19%
Query Match:	10.85%
DB:	10
	2
Length:	120
Matches:	92
Conservative:	2
Mismatches:	4
Indels:	10
Gaps:	2

US-09-502-945-4 (1-2236) x US-09-925-299-778 (1-120)

QY	1315	GGAGTGGAACTCGAGTCCGAGCCCGCAGATGACCTGGATGGAGGCACGGAGGACGAGGA	1374
Db	23	GlyLeuArgProValLeuGlnPro-----ArgGlnGly	33
QY	1375	GAGCAGAGATTCCGGAAATATAGGAAGGCTTTGACCCTACTCTATGCTTCACCCACAG	1434
Db	34	---GlnAspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGlu	52
QY	1435	CAGATCATGGGAAGGATGTCGGCTCTACCCATCAGAAGGAGGGATCCCTTAGACCTG	1494
Db	53	GlnIleMetGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeuAspLeu	72
QY	1495	GCCTTGGAAAGCGGTGTGGACTCCCCCATTTGGGAAGGTGGTCTTCTGCTGTATTGAG	1554
Db	73	AlaLeuGluGlyGlyValAspSer***IleGlyLysValValSerAlaValTyrGlu	92
QY	1555	CGGGGAGCTGCTGAGCGCATGGTGCATTTGCAAGGGGACGAGATCATGGCAATCAAC	1614
Db	93	ArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsn	112
QY	1615	GCCAAGATTGTGACAGACTACACC	1638
Db	113	GlyLysIleValThrAspTyrThr	120

```

RESULT 4
US-09-739-907-60
; Sequence 60, Application US/09739907
; Patent No. US2001001289A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (98)
; OTHER INFORMATION: xaa equals stop translation
US-09-739-907-60

```

Alignment Scores:

Pred. No.:	2.08e-23	Length:	98
Score:	417.00	Matches:	88
Percent Similarity:	83.81%	Conservative:	0
Best Local Similarity:	83.81%	Mismatches:	17
Query Match:	10.12%	Indels:	1
DB:	10	Gaps:	1

US-09-502-945-4 (1-2236) x US-09-739-907-60 (1-98)

Qy	676	GGCGTCGAGCGACCGCTGGGCTCCCTGGAAATCGGAAACACAGAGAGAAAGTGCTTTC	735
Db	10	GlyValArgGlySerLeuGlySerProGlyAsnArgGluAsnLysGluLysValPhe	29
Qy	736	ATCAGCCTGCTAGGCTCCCGAGCGCTTGCTGCAGCATTTCCAGCGCGCCCATCCAGAAG	795
Db	30	IleSerLeuValGlySerArgGlyLeuGlyCysSerIleSerSerGlyProIleGlnLys	49
Qy	796	CTGCGCATCTTTATCAGCCATGTAAACCTGGCTCCTCTCTCTCTGAGTGGGATTGGAG	855
Db	50	ProGlyIlePheIleSerHisValLysProGlySerLeuSerAlaGluValGlyLeuGlu	69
Qy	856	ATAGGGGACACAGATTGTCGAAGTCAATGGCGTGCACCTTCTAACTGGATCACAAGGAG	915
Db	70	IleGlyAspGlnIleValGluValAsnGlyValAspPheSerAsnLeuAspHisLysGlu	89
Qy	916	GCTGTTAAATGCTGTAARAATACCCGACGCTGACCATCTCCATTCTAGCTGCAGCTGGC	975
Db	90	-Leu-	 -----GlnLeuAla 93
Qy	976	CGGAGCTGTTC	988
Db	93	aglySerCysSer	97

RESULT 5

```

; Sequence 119, Application US/09739907
; Patent No. US2001012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secretes
; FILE REFERENCE: P2022p1
; CURRENT APPLICATION NUMBER: US/09/73
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ. ID NO 119
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-7399-907-119

```

Alignment Scores:

Alignment Scores:			
Pred. No.:	2, 21e-23	Length:	113
Score:	417.00	Matches:	88
Percent Similarity:	83.81%	Conservative:	0
Best Local Similarity:	83.81%	Mismatches:	0
Query Match:	10.12%	Indels:	17
DB:	10	Gaps:	1

US-09-502-945-4 (1-2236) x US-09-739-907-119 (1-113)

Qy 676 GCGTGGAGGCAGCCTGGGCTCCCTGGAAATCGGAAACAGGAGAAGGCTTC 735


```
Db 26 GlyValArgGlySerLeuGlySerProGlyAsnArgGluAsnLysGluLysLysValPhe 45
QY 736 ATGAGCCCTGTAGCTCCCGAGGCTTGGCTGAGAGATTTCCAGCGGCCCATCCAGAG 795
Db 46 IleSerLeuValGlySerArgGlyLeuGlycSerIleSerGlyProIleGlnLys 65
QY 796 CCTGGCATCTTTATCAGCATGTGAACCTGGCTGCCCTGTCTGTGAGGTGGGATTGGAG 855
Db 66 ProGlyIlePheIleSerHisValLysProGlySerLeuSerAlaGluValGlyLeuGlu 85
QY 856 ATAGGGGACCAAGATTGCGAAGTCAATGGCGTGCAGCTTCTTAACCTGGATCATCAAGAG 915
Db 86 IleGlyAspGlnIleValGluValAsnGlyValAspPheSerAsnLeuAspHisLysGlu 105
QY 916 GCTGTAATGTGCTGAAAAATAGCGCAGCGCTGACCATCTCCATTGTAGTGCAGGTGGC 975
Db 106 -Leu-----GlnLeuAl 109
QY 976 CGGGAGCTGTTCA 988
Db 109 aglyserCysSer 113
RESULT 6
US-09-739-907-111
; Sequence 111, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09739,907
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 111
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-907-111
Alignment Scores:
Pred. No.: 1,37e-22 Length: 95
Score: 406.00 Matches: 90
Percent Similarity: 60.13% Conservative: 2
Best Local Similarity: 58.82% Mismatches: 1
Query Match: 9.85% Indels: 60
DB: 10 Gaps: 1
US-09-502-945-4 (1-2236) x US-09-739-907-111 (1-95)
QY 1752 GCTGAAGTCCAAAGGGGAAACCAATTCACGCGTTAGGAAACAGTACGCTCCGGCCCA 1811
Db 3 SerGluValGlnLysGlyLysProAsnSer-ProLeuGlyAsnSerGluLeuArgProHi 22
QY 1812 CTCCTGTGAACAAAGCCCTCGACAGCTTTCAGAGAGCCCATGACACACACCATG 1871
Db 22 IleuValAsnThrLysProArgThrSerLeuGluArgGlyHis----- 36
QY 1872 GCATCCTTGGGACCTGAATCTATCACCAGGAATCTCAAACCTCCCTTTGGCCCTGAACCA 1931
Db 36 ----- 36
QY 1932 GGGCCAGATAAGGAACAGCTCGGCCACTTTTTTTGAAGGCCAAATGCTGGAGGAAGGGAGC 1991
```

```
Db 36 ----- 36
QY 1992 AGCCAGCCGTTTGGGAGAGAATCTCAAGGATCCAGACTCTCATTCCTTCTCTGGCCCA 2051
Db 37 -----Thr-IleProPheLeuIrrProS 44
QY 2052 GTGAATTTGGTCTCTCCAGCTTTGGGGACTCCTTCTTGAACCCCTAATAAGACCCAC 2111
Db 44 erGluPheGlyLeuSerGlnLeuIrrPglyThrProSerLeuAsnProAsnLysThrProL 64
QY 2112 TGGAGTCTCTCTCTCCATFCCCTCTCTCTGCCCTCTGCTTAATTCGTCGACGATTG 2171
Db 64 euGluSerLeuSerLeuHisProSerProLeuProSerAlaLeuIleAlaAlaArgIleV 84
QY 2172 TCACTCCAAACCTTACTCTGAGCTCATTAATAAA 2206
Db 84 alThrProAsnLeuThrLeuSerSerLeuIleLys 95
RESULT 7
US-10-078-090-151
; Sequence 151, Application US/10078090
; Publication No. US2003004815A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karra, Kalpana
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and
; FILE REFERENCE: DEX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 151
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-078-090-151
Alignment Scores:
Pred. No.: 7,91e-16 Length: 260
Score: 318.50 Matches: 85
Percent Similarity: 52.32% Conservative: 39
Best Local Similarity: 35.86% Mismatches: 84
Query Match: 7.73% Indels: 29
DB: 9 Gaps: 6
US-09-502-945-4 (1-2236) x US-10-078-090-151 (1-260)
QY 358 GTGCGTCTGGACCGCTGTCACCCGAGAGCGCTGAGTGTGCTGTGCTGTGCTGGAG 417
Db 1 LeuArgArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleArgGlyGlySerGlu 19
QY 418 TTGGGTGTGGCTTCTTCTCCACTCTCCACTCATCAAGGCGGTGAGGACAGACAGCGTGG 477
Db 20 HisGlyValGlyIleTyrrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
QY 478 CTCAGGTAGGGACGAGATCGTCGCGATCAATGATATTCATCTCTCTCTACCCAT 537
Db 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
QY 538 GAGGAGGTGATCAACCTTCATTGCAACCAAGAACTGTGTCTCAAGTGAAGTGAACATC 597
Db 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrrSerAla 79
QY 598 GGCCTGATCCCGGTGAAGAAAGCTCTCTGATGAGCCCTCTACTTGGCAGTATGTGGAT--- 654
```

```

Db 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAspPro 97
    ||| ||||| :||| ||||| |||||
QY 655 -----CAGTTTGTGCGGAATCTGGGCGCTCGAGGAGCGCTGGC----- 596
    :||| :||| :|||
Db 98 GlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg 117
    :||| :||| :|||
QY 697 ---TCCCTCGGAATCGGAAC-----AAGGAGAAGAAG 729
    |||:|||| :|||
Db 118 GlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGlnGlyLys 137
    |||:|||| :|||
QY 730 GTCCTTCATCAGCTGTAGCTCCCGAGGCTTGGCTGCAGCATTTCCAGCGCCCATC 789
    ||| :||| :|||
Db 138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyAlaGlu 157
    ||| :||| :|||
QY 790 CAGAACCTCGGCATTTATCCCATGCAACCTGCAACCTGCTCTCTCTCTCTCTCTCT 849
    |||:|||| :|||
Db 158 TyrGlyLeuGlyIleThrGlyValAspProGlySerGluAlaGlySerGly 177
    |||:|||| :|||
QY 850 TTGGAGATAGGGAGCAGCATTTGCAAGTCAATGGCGTCTCTCTCTCTCTCTCTCTCT 909
    |||:|||| :|||
Db 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
    |||:|||| :|||
QY 910 AAGGAGCTGTAAATGCTGTAATAATACCGCAGCCTGACCATCTCCATTGTAGCTGCA 969
    |||:|||| :|||
Db 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThrVal----- 214
    |||:|||| :|||
QY 970 GCTGGCGGGAGCTGTTTCATGACAGACCGGGAGCGCTGGCAGAGCGGGG 1020
    |||:|||| :|||
Db 215 -----LysaspValGlyArgLeuProHisAlaArg 224
    |||:|||| :|||

RESULT 8
US-09-951-402-3
; Sequence 3, Application US/09951402
; Patent No. US20020168752A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/951,402
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 09/306,998
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 60/084,740
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; TYPE: PRT
; LENGTH: 2037
; ORGANISM: Homo sapiens
US-09-951-402-3

Alignment Scores:
Pred. No.: 1,29e-11 Length: 2037
Score: 267.50 Matches: 142
Percent Similarity: 35.81% Conservative: 109
Best Local Similarity: 20.26% Mismatches: 223
Query Match: 6.49% Indels: 228
DB: 9 Gaps: 26

US-09-502-945-4 (1-2236) x US-09-951-402-3 (1-2037)
QY 388 CTCGCCCTGAGTGTGGTGGCTGTGGCTTTCATCTCCACCTC 447
    |||||:|||||
Db 1018 LeuGlyMetThrValSerAlaAsnLysAsp---GlyLeuGlyMetIleValArgSerIle 1036
    |||:|||| :|||
QY 448 ATCAAGCGGTCCAGCAGACAGCTCGG---CTCCAGGTAGGGAGCAGATCGTCGG 504
    |||:|||| :|||
Db 1037 IleHisGlyGlyAlaIleSerArgaspGlyArgIleAlaIleGlyAspCysIleLeuSer 1056
    |||:|||| :|||
QY 505 ATCAATGATATTCCATCTCTCTCTATCCCATGAGGAGGTTCATCACTTCGTAACC 564
    |||:|||| :|||

```

```

Db 1057 IleAsnGluGluSerThrIleSerValThrAsnAlaGlnAlaArgAlaMetLeuArg--- 1075
    ||||| ||| ||| |||:|||| :|||
QY 565 AAGAAACATGTGTCCATCAAAAGTGAGACACATCGGCCCTGATCCCGCTGAAAAAGCTCTCCT 624
    ||||| ||||| |||||
Db 1076 -----ArgHis---SerLeuIle-----GlyPro 1082
    :||| :||| :|||
QY 625 GATGAGCCCTCACCATTGG-----CAGTATGTGGATCAGTTTGTGTCTCGAATCTCGG 675
    ||| :|||:|||| :|||
Db 1083 AspIleLysIleThrTyrValProAlaGluHisLeuGluGluPheLysIleSerLeuGly 1102
    ||||| :|||:|||| :|||
QY 676 GCGGTGCGAGGCGAGCTG----- 693
    |||:|||| :|||
Db 1103 GlnGlnSerGlyArgValMetAlaLeuAspIlePheSerSerTyrThrGlyArgAspIle 1122
    |||:|||| :|||
QY 694 -----GGCTCCCTCGAAATCGGAAAAACAAGGAG----- 723
    ||| |||
Db 1123 ProGluLeuProGluArgGluGluGlyGluGluSerGluLeuGlnAsnThrAla 1142
    ||| |||
QY 724 -----AAGAGGTC 732
    ||| |||
Db 1143 TyrSerAsnTrpAsnGlnProArgArgValGluLeuTrpArgGluProSerLysSerLeu 1162
    ||| |||
QY 733 TTCATCAGCCTGGTAGGCTCCCGAGGCTTGGCTGCAGCATTTCCAGCGGCCCATCCAG 792
    |||||:||||| :|||
Db 1163 GlyIleSerIleValGlyArgGlyMetGlySerArgLeuSerAsnGlyValuMet 1182
    |||||:||||| :|||
QY 793 AAGCCTGTCATCTTTATCAGCCATGTGAACCTGGCTCCCTCTCTCTCTCTCTCTCTCT 849
    |||:||||| :|||
Db 1183 Arg---GlyIlePheIleLysHisValLeuGluAspSerProAlaGlyLysAsnGlyThr 1201
    |||:||||| :|||
QY 850 TTGGAGATAGGGAGCAGATTGTCGAAGTCAATGGCTGCGACTTCTTAACCTGGATCAC 909
    |||||:|||||:||||| :|||
Db 1202 LeuLysProGlyAspArgIleValGluValAspGlyMetAspLeuArgAspAlaSerHis 1221
    |||||:|||||:||||| :|||
QY 910 AAGGAGCTGTAAATGCTGCTGAAATAATAGCCGAGCCCTGACCATCTCCATTGTAGCTGCA 969
    |||||:|||||:||||| :|||
Db 1222 GluGlnAlaValGluAlaIleArgLysAlaGlyAsnProValValPheMetValGlnSer 1241
    |||||:|||||:||||| :|||
QY 970 GCTGGCGGGAGCTTTCATGACAGACCGGAGCGGCTGGCAGAGCGCGGCGGCGTGGAG 1029
    |||:||||| :|||
Db 1242 -----IleIleAsnArgProArgAlaProSerGlnSerGluSerGlu 1255
    |||:||||| :|||
QY 1030 CTGAGCGGCGAGGAGCTTCTCATCGAAGCGGCTGGCGATGGAGTCCAACAAGATCCTC 1089
    |||||:||||| :|||
Db 1256 ProGluLysAlaProLeuCys----- 1262
    |||:||||| :|||
QY 1090 CAGGAGCAGCAGGAGATGGAGCGCGCAAGAGAAAGAAATTTGCCAGAGGCGAGCAGAG 1149
    |||:||||| :|||
Db 1263 -----SerValProProProProSerAlaPheAlaGluMetGlySerAsp 1278
    |||:||||| :|||
QY 1150 GAAATGAGAGATACCGGAGGAGATGGAACAGATTGTAGAGGAGGAGAGAGAGTTTAAG 1209
    |||:||||| :|||
Db 1279 HisThrGlnSerSerAlaSerLysIleSerGlnAspValAspLysGluAspGluPheGly 1298
    |||:||||| :|||
QY 1210 AAGCAATGGGAA-----GAAGACTGGGCTCAAAGGAACAGCTACTCTTTCCTCTCTCTCT 1260
    |||||:||||| :|||
Db 1299 TyrSerTrpLysAsnIleArgGluArgTyrGly----- 1309
    |||||:||||| :|||
QY 1261 ACCATCATCTGTGAGGTACACCCAGTACCCCTCTCCGCAAG----- 1299
    |||||:||||| :|||
Db 1310 ThrLeuThrGlyGluLeuHisMetIleGluGlyHisSerGlyLeuGlyLeu 1329
    |||||:||||| :|||
QY 1300 -----CCA 1302
    |||:||||| :|||
Db 1330 SerLeuAlaGlyAsnLysAspArgSerArgMetSerValPheIleValGlyIleAspPro 1349
    |||:||||| :|||
QY 1303 AAGTATGATCAGGAGGTGGAACTCGAGTCCAGCCCGCAGATGACCTG-----GAT 1353
    |||||:||||| :|||
Db 1350 AsnGlyAlaAlaGlyLysAspGlyArgLeuGlnIleAlaAspGluLeuLeuGluIleAsn 1369
    |||||:||||| :|||
QY 1354 GGA-----GGCAGCGGAGGAGCAGGAGCAGGAGATTTCCTCGGAAA----- 1392
    |||:||||| :|||

```

Db 1370 GlyGlnIleLeuTyrGlyArgSerHisGlnAsnAlaSerSerIleIleLysCysAlapro 1389
Qy 1393 -----TATGAGGAGGCTTTGAC 1410
Db 1390 SerLysValLysIleIlePheIleArgAsnLysAspAlaValAsnGlnMetAlaValCys 1409
Qy 1411 CCTACTCTATGTCACCCCA-----GAGCAGATCATGGGAGAGGATGTC 1455
Db 1410 ProGlyAsnAlaValGluProLeuProSerAsnSerGluAsnLeuGlnAsnLysGluThr 1429
Qy 1456 CGGCTCTACGATCAAGAAGAGGATTCCTAGACCTGGCCCTG----- 1500
Db 1430 GluProThrValThrSerAspAlaAlaValAspLeuSerSerPheLysAsnValGln 1449
Qy 1501 -----GAGGCGGTCTG-----GACTCCCCC 1521
Db 1450 HisLeuGluLeuProLysAspGlnGlyLeuGlyIleAlaIleSerGluGluAspThr 1469
Qy 1522 ATTGGAAGGTGGTCTGCTGCTGATGAGCGGGGAGCTGCTGAGCGCATGGTGGC 1581
Db 1470 LeuSerGlyValIleIleLysSerLeuThrGluHisGlyValAlaAlaThrAspGlyArg 1489
Qy 1582 ATTGGAAGGGGAGGATCATGCAATCAAGCGCAAGATGTGACAGACTACACCTG 1641
Db 1490 LeuLysValGlyAspGlnIleLeuAlaValAspAspGluIleValValGlyTyrProIle 1509
Qy 1642 GCTGAGGCTGAGCTGCTGCGCAGAAAGCGCTGGAATCAGGCGGGGAGCTGCACCTT 1701
Db 1510 GluLysPheIleSerLeuLeuLysThrAla----- 1519
Qy 1702 GTGGTTGCGCTGCTGCCCCCAAGGAGTATGACGATGAGCTGACCTTCTTGTGTA--GTC 1760
Db 1520 -----LysMetThrValLysLeuThrIle 1527
Qy 1761 CAAGAAGGGAAACCAATTCACGCGTTAGGAAACAGTAGTCCGCGGCCACCTCGTGAA 1820
Db 1528 HisAlaGluAsnProAspSerGlnAlaValProSerAlaAlaGlyAlaAlaSerGlyGlu 1547
Qy 1821 CACAAGACCTCGGACGCTTGAGAGGCGCACATGACACACACATGAGTCCCTCTG 1880
Db 1548 LysLysAsnSerSerGlnSerLeuMetValPro-----GlnSerGlySerPro 1563
Qy 1881 GGACCTGAATCTATC-----ACCAGGAATCTCAACTCC-----TTTGGCCCTGAA 1928
Db 1564 GluProGluSerIleArgAsnThrSerArgSerThrProAlaIlePheAlaSerAsp 1583
Qy 1929 CCA----- 1931
Db 1584 ProAlaThrCysProIleIleProGlyCysGluThrThrIleGluIleSerLysGlyArg 1603
Qy 1932 -----GGGCCAGATAAGGAACAGCTCGGGCCACTTTTGA 1967
Db 1604 ThrGlyLeuGlyLeuSerIleValGlyGlySerAspThrLeuLeuGlyAlaIleIleIle 1623
Qy 1968 AGGCCAATGTGGAGGAAGGAGCAGCCAGCGCTTTGGGAGAGATCTCAAGGATCCAGA 2027
Db 1624 HisGluValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaGlyaspGln 1643
Qy 2028 CTC 2030
Db 1644 Ile 1644

RESULT 9

US-09-951-401-3
; Sequence 3, Application US/09951401
; Patent No. US20020115104A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/951,401
; CURRENT FILING DATE: 2001-09-14

; PRIOR APPLICATION NUMBER: US 09/306,998
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 60/084,740
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-951-401-3

Alignment Scores: 1.29e-11 Length: 2037
Pred. No.: 267.50 Matches: 142
Score: 35.81% Conservative: 109
Best Local Similarity: 20.26% Mismatches: 223
Query Match: 6.49% Indels: 228
DB: 10 Gaps: 26

US-09-502-945-4 (1-2236) x US-09-951-401-3 (1-2037)

Qy 388 CTGGGCTGAGTGTGCTGGTGGCTGAGCTTTGGCTGCTTCATCTCCACCTC 447
Db 1018 LeuGlyMetThrValSerAlaAsnLysAsp---GlyLeuGlyMetIleValArgSerIle 1036
Qy 448 ATCAAAGCGGTGACGAGCAGACAGCGTCCGG---CTCCAGGTAGGGACGAGATCGTCCGG 504
Db 1037 IleHisGlyGlyAlaIleSerArgAspGlyArgIleAlaIleGlyAspCysIleLeuSer 1056
Qy 505 ATCAATGGATATTCATCTCTCTGTACCATGAGGAGTGCATCAACCTCATTCGAAACC 564
Db 1057 IleAsnGluGluSerThrIleSerValThrAsnAlaAlaArgAlaMetLeuArg--- 1075
Qy 565 AAGAAACTGTGTCCATCAAGTGAACATCGGCTGATCCCGTGAAGAAAGCTCTCCT 624
Db 1076 -----ArgHis---SerLeuIle-----GlyPro 1082
Qy 625 GATGAGCCCTCACTTGG-----CAGTATGTGGATCAGTTTGTGTCGGAATCTGGG 675
Db 1083 AspIleLysIleThrValProAlaGluHisLeuGluGluPheLysIleSerLeuGly 1102
Qy 676 GGCTGGGAGCAGCCCTG----- 693
Db 1103 GlnGlnSerGlyArgValMetAlaLeuAspIlePheSerSerTyrThrGlyArgAspIle 1122
Qy 694 -----GGCTCCCTGGAATCGGGGAAACAAGAG----- 723
Db 1123 ProGluLeuProGluArgGluGluGlyGluGluSerGluLeuGlnAsnThrAla 1142
Qy 724 -----AAGAAGTCT 732
Db 1143 TyrSerAsnTrpAsnGlnProArgArgValGluLeuTrpArgGluProSerLysSerLeu 1162
Qy 733 TTCTACAGCTGGTAGCTCCCGAGGCTTGGTGCAGCATTTCCAGCGGCCCATCCAG 792
Db 1163 GlyIleSerIleValGlyArgGlyMetGlySerArgLeuSerAsnGlyGluValMet 1182
Qy 793 AAGCCTGGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGCTGCTGAGTGGGA--- 849
Db 1183 Arg---GlyIlePheIleLysHisValLeuGluAspSerProAlaGlyLysAsnGlyThr 1201
Qy 850 TTGGAGATAGGGGACCATGTGCAAGTCAATGCGCTCGACTTCTTAACCTGGATCAC 909
Db 1202 LeuLysProGlyAspArgIleValGluValAspGlyMetAspLeuArgAspAlaSerHis 1221
Qy 910 AAGGAGGCTGTAATGTGCTGAAATAATAGCGCAGCTGACCATCTCCATGTAGCTCCA 969
Db 1222 GluGlnAlaValGluAlaIleArgLysAlaGlyAsnProValPheMetValGlnSer 1241
Qy 970 CTGCGCGGGAGCTGTTTCATACAGACCGGAGCGGCTGGCAGAGCGCGGACCGGTGAG 1029
Db 1242 -----IleIleAsnArgProArgAlaProSerGlnSerGluSerGlu 1255

```
QY 1030 CTGACGGCGCAGGAGCTTCTCATGCAGAGAGCGGCTGGCGATGGAGTCCCAACAGATCCTC 1089
Db 1256 ProGluLysAlaProLeuCys----- 1262
QY 1090 CAGGACGACGAGGATGGAGCGGCAAGAGAAATTTGCCAGAGGACGACAGAG 1149
Db 1263 -----SerValProProProSerAlaPheAlaGluMetGlySerAsp 1278
QY 1150 GAAATGAGACATCCGGAAGGAGATGGACAGATTTAGAGAGGAGAAAGATTTTAAAG 1209
Db 1279 HisThrGlnSerSerAlaSerLysIleSerGlnAspValAspLysGluAspGluPheGly 1298
QY 1210 AAGCAATGGGAA-----CAAGACTGGGCTCAAGGACAGCTACTCTTGCTCTAAA 1260
Db 1299 TyrSerTrpLysAsnIleArgGluArgTyrGly----- 1309
QY 1261 ACCATCACTGCTGAGTACACCCAGTACCTCCCTCCGAAG----- 1299
Db 1310 ThrLeuThrGlyGluLeuHisMetIleGluLeuGlyLysGlyHisSerGlyLeuGlyLeu 1329
QY 1300 -----CCCA 1302
Db 1330 SerLeuAlaGlyAsnLysAspArgSerArgMetSerValPheIleValGlyIleAspPro 1349
QY 1303 AAGTATGATCAGGAGTGGACCTGAGCTCGAGCCGCGAGATGACCTG-----GAT 1353
Db 1350 AsnGlyAlaAlaGlyLysAspGlyArgLeuGlnIleAlaAspGluLeuLeuGluIleAsn 1369
QY 1354 GGA-----GGCAGGAGGAGGAGGAGGAGGAGGATTTCCGGAAA----- 1392
Db 1370 GlyGlnIleLeuTyrGlyArgSerHisGlnAsnAlaSerSerIleLeuLysCysAlaPro 1389
QY 1393 -----TATGAGGAAGGCTTTGAC 1410
Db 1390 SerLysValLysIleIlePheIleArgAsnLysAspAlaValAsnGlnMetAlaValCys 1409
QY 1411 CCTTACTCTATGTTCCACCCCA-----GAGCAGATCATCGGGAAGGATGTC 1455
Db 1410 ProGlyAsnAlaValGluProLeuProSerAsnSerGluAsnLeuGlnAsnLysGluThr 1429
QY 1456 CGGCTCTCTACGATCAAGAGGAGGAGGATCTTAGACCTGGCGCTG----- 1500
Db 1430 GluProThrValThrThrSerAspAlaAlaValAspLeuSerPheLysAsnValGln 1449
QY 1501 -----GAGGCGGTGTG-----GACTCCCCC 1521
Db 1450 HisLeuGluLeuProLysAspGlnGlyLeuGlyIleAlaIleSerGluGluAspThr 1469
QY 1522 ATTGGGAAGGTGGTCTTCTGCTGTGTATGAGCGGGGAGCTGCTGACGGCATGGTGGC 1581
Db 1470 LeuSerGlyValIleIleLysSerLeuThrGluHisGlyValAlaAlaThrAspGlyArg 1489
QY 1582 ATTGTGAAAGGGACGAGATCATGCAATCAACGCAAGATTTGACAGACTACACCTG 1641
Db 1490 LeuLysValGlyAspGlnIleLeuAlaValAspAspGluIleValValGlyTyrProIle 1509
QY 1642 GCTGAGGCTGAGCTGCTGCTGAGAGGCGCTGGAATCAGGCGGGGAGTGCATCGACCTT 1701
Db 1510 GluLysPheIleSerLeuLeuLysThrAla----- 1519
QY 1702 GTGGTTGCGCTGCTGCCCCCAAGAGGATATCAGATGAGCTGACCTTCTTCTGAA-GTC 1760
Db 1520 -----LysMetThrValLysLeuThrIle 1527
QY 1761 CAAAGGGGAAACCAATTACGGCTTAGGAAACAGTACGCTCGCGCCACCTCGTCAA 1820
Db 1528 HisAlaGluAsnProAspSerGlnAlaValProSerAlaAlaGlyAlaAlaSerGlyGlu 1547
QY 1821 CACAAAGCTCGGACCGCTTGTAGAGAGGCGCATGACACACACAGATGGCATCTTGT 1880
Db 1548 LysLysAsnSerSerGlnSerLeuMetValPro-----GlnSerGlySerPro 1563
```

```
QY 1381 GGACCTGAATCTATC-----ACCCAGGAATCTCAAACTCCC-----TTTGGCCCTGAA 1928
Db 1564 GluProGluSerIleArgAsnThrSerArgSerThrProAlaIlePheAlaSerAsp 1583
QY 1929 CCA----- 1931
Db 1584 ProAlaThrCysProIleIleProGlyCysGluThrThrIleGluIleSerLysGlyArg 1603
QY 1932 -----GGCCAGATAAAGAACAGCTCGGGCCACTTTTTTGA 1967
Db 1604 ThrGlyLeuGlyLeuSerIleValGlyGlySerAspThrLeuLeuGlyAlaIleIle 1623
QY 1968 AGGCAATGTGGAGGAAAGGAGCAGCAGCCGTTTGGGAGAGATCTCAAGATCCAGA 2027
Db 1624 HisGluValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaGlyAspGln 1643
QY 2028 CTC 2030
Db 1644 Ile 1644
```

RESULT 10

```
US-09-922-101-3
; Sequence 3, Application US/09922101
; Patent No. US2002014671A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/922,101
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/306,998
; PRIOR FILING DATE: 1999-05-07
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-101-3
```

Alignment Scores:

```
Pred. No.: 1,29e-11 Length: 2037
Score: 267.50 Matches: 142
Percent Similarity: 35.81% Conservative: 109
Best Local Similarity: 20.26% Mismatches: 223
Query Match: 6.49% Indels: 228
DB: 10 Gaps: 26
```

US-09-502-945-4 (1-2236) x US-09-922-101-3 (1-2037)

```
QY 388 CTGGCGCTGAGTGTGCTGGCTGGAGTTGGCTGTGGCTCTTCACTCCACCTC 447
Db 1018 LeuGlyMetThrValSerAlaAsnLysAsp--GlyLeuGlyMetIleValArgSerIle 1036
QY 448 ATCAAGGCGGTGAGGAGCAGACGCTCGGG---CTCAGGTAGGGACGAGATGTCGCGG 504
Db 1037 IleHisGlyGlyAlaIleSerArgAspGlyArgIleAlaIleGlyAspCysIleLeuSer 1056
QY 505 ATCAATGGATATCCATCTCTCTGTACCCATCAGGAGGTCATCAACCTCATTCGNAACC 564
Db 1057 IleAsnGluSerThrIleSerValThrAsnAlaGlnAlaArgAlaMetLeuArg--- 1075
QY 565 AAGAAACTGTGTCCATCAAGTGAGACATCGGCTGATCCCGGTGAAAGCTCTCCT 624
Db 1076 -----ArgHis---SerLeulle-----GlyPro 1082
QY 625 GATGAGCCCTCACTTGG-----CAGTATGTGGATCAGTTGTGTGCGAATCTGG 675
Db 1083 AspIleLysIleThrTyrValProAlaGluHisLeuGluGluPheLysIleSerLeuGly 1102
QY 676 GCGTGCAGGACCCCTG----- 693
```

Db 1103 GlnGlnSerGlyArgValMetAlaLeuAspIlePheSerSerTyrThrGlyArgAspIle 1122
Qy 694 -----GGCTCCCTCGGAATCGGAAACAAAGAG----- 723
Db 1123 ProGluLeuProGluArgGluGluGlyGluGluSerGluLeuGlnAsnThrAla 1142
Qy 724 -----AAGAAGTGC 732
Db 1143 TyrSerAsnTrpAsnGlnProArgArgValGluLeuTrpArgGluProSerLysSerLeu 1162
Qy 733 TTCATCAGCCCTGGTAGCTCCGAGCCCTGGCTGCACATTCACAGCGGCCCATCCAG 792
Db 1163 GlyIleSerIleValGlyArgGlyMetGlySerArgLeuSerAsnGlycyluValMet 1182
Qy 793 AAGCCTGGCATCTTATCAGCCCTCGAAACCTCGCTGCTGCTGAGTGGGA--- 849
Db 1183 Arg---GlyIlePheIleLysHisValLeuGluAspSerProAlaGlyLysAsnGlyThr 1201
Qy 850 TTGGAGATAGGGACGACGATTGTGAAGTCAATGGCGTGCATCTCTPAACCTGGATCAC 909
Db 1202 LeuLysProGlyAspArgIleValGluValAspGlyMetAspLeuArgAspAlaSerHis 1221
Qy 910 AAGGAGGCTGTAATGCTGTAAGAAATACCCGACGCTGACCATCTCCATTGTAGCTGCA 969
Db 1222 GluGlnAlaValGluAlaIleArgLysAlaGlyAsnProValPheMetValGlnSer 1241
Qy 970 GCTGGCGGGAGCTGTTCATGACAGACGGGAGCGGCTGGCAGAGCGCGGCGCTGAG 1029
Db 1242 -----IleIleAsnArgProArgAlaProSerGlnSerGluSerGlu 1255
Qy 1030 CTGACAGCGCAGAGCTTCTCATGAGAGCGGCTGGCGATGGATGCCAACAGATCTC 1089
Db 1256 ProGluLysAlaProLeuCys----- 1262
Qy 1090 CAGGAGCAGCAGAGATGAGCGGCAAGGAGAAAGAAATGCCAGAGGCCAGCAGAG 1149
Db 1263 -----SerValProProProProSerAlaPheAlaGluMetGlySerAsp 1278
Qy 1150 GAAATATGAGATACCGGAAGGAGATGGAACAGATTGTAGAGGAGGAAGATTAAAG 1209
Db 1279 HisThrGlnSerSerAlaSerLysIleSerGlnAspValAspLysGluAspIleGly 1298
Qy 1210 AAGCAATGGGA-----GAAGACTGGGCTCAAGGAACAGCTACTCTGTGCTAAA 1260
Db 1299 TyrSerTrpLysAsnIleArgGluArgTyrGly----- 1309
Qy 1261 ACCATCATCTGCTGAGTACACCCAGTACCCCTCGCAAG----- 1299
Db 1310 ThrLeuThrGlyGluLeuHisMetIleGluLeuGluLysGlyHisSerGlyLeuGlyLeu 1329
Qy 1300 -----CCA 1302
Db 1330 SerLeuAlaGlyAsnLysAspArgSerArgMetSerValPheIleValGlyIleAspPro 1349
Qy 1303 AAGTATGATCAGGAGTGGAACTGAGCTCGAGCCCGCAGATGACCTG-----GAT 1353
Db 1350 AsnGlyAlaAlaGlyLysAspGlyArgLeuGluGlnIleAlaAspGluLeuGluIleAsn 1369
Qy 1354 GGA-----GGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1392
Db 1370 GlyGlnIleLeuTyrGlyArgSerHisGlnAsnAlaSerSerIleIleLysCysAlaPro 1389
Qy 1393 -----TATCAGGAAGGCTTTGAC 1410
Db 1390 SerLysValLysIleIlePheIleArgAsnLysAspAlaValAsnGlnMetAlaValCys 1409
Qy 1411 CCTACTCTATGTCACCCCA-----GACAGATCATGGGGAAGGATGTC 1455
Db 1410 ProGlyAsnAlaValGluProLeuProSerAsnSerGluAsnLeuGlnAsnLysGluThr 1429
Qy 1456 CGGCTCTAGCATCAAGAGGAGGATCTTAGACCTGGCCCTG----- 1500
Db 1430 GluProThrValThrSerAspAlaAlaValAspLeuSerSerPheLysAsnValGln 1449

Qy 1501 -----GAAGCGGTGTG-----GACTCCCC 1521
Db 1450 HisLeuGluLeuProLysAspGlnGlyLeuGlyIleAlaIleSerGluGluAspThr 1469
Qy 1522 ATTGGGAAGTGTCTGCTGCTGCTATGACGGGGAGCTGCTGAGCGGCATGGTGC 1581
Db 1470 LeuSerGlyValIleIleLysSerLeuThrGluHisGlyValAlaAlaThrAspGlyArg 1489
Qy 1582 ATTGTGAAGGGGACGAGATCATGCAATCAACGGCAAGATTGTGACAGACTACACCTG 1641
Db 1490 LeuLysValGlyAspGlnIleLeuAlaValAspGluIleValGlyTyrProIle 1509
Qy 1642 GCTGAGCTGACGCTGCCCTGCACAGAGCCCTGGAATCAGGCGGGGACTGGATCGACCTT 1701
Db 1510 GluLysPheIleSerLeuLeuLysThrAla----- 1519
Qy 1702 GTGTTGCTGCTGCTCCCGCCCAAGGAGTATGACGATGAGCTGACCTTCTTGTCTGAA-GTC 1760
Db 1520 -----LysMetThrValLysLeuThrIle 1527
Qy 1761 CAAAAGGGGAAACCAAAATTCACGCGTTAGGAAACAGTGAGCTCGCGCCACCTCGTGAA 1820
Db 1528 HisAlaGluAsnProAspSerGlnAlaValProSerAlaAlaGlyAlaAlaSerGlyGlu 1547
Qy 1821 CACAAAGCCTCGGACCGCTTGAGAGAGGCCACATGACACACAGATGGCATCTTGT 1880
Db 1548 LysLysAsnSerSerGlnSerLeuMetValPro-----GlnSerGlySerPro 1563
Qy 1881 GGACTCAATCTATC-----ACCAGGAATCTCAAACTCCC-----TTTGGCCCTGAA 1928
Db 1564 GluProGluSerIleArgAsnThrSerArgSerThrProAlaIlePheAlaSerAsp 1583
Qy 1929 CCA----- 1931
Db 1584 ProAlaThrCysProIleIleProGlyCysGluThrThrIleGluIleSerLysGlyArg 1603
Qy 1932 -----GGCCAGATAAGGAACAGCTCGGCGCCACTTTTGTGA 1967
Db 1604 ThrGlyLeuGlyLeuSerIleValGlyGlySerAspThrLeuLeuGlyAlaIleIle 1623
Qy 1968 AGGCAATGTGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2027
Db 1624 HisGluValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaGlyAspGln 1643
Qy 2028 CTC 2030
Db 1644 Ile 1644

RESULT 11

US-09-919-497-59
; Sequence 59, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59
; LENGTH: 767
; TYPE: PRT
; ORGANISM: Homo sapiens
us-09-919-497-59

Alignment Scores: 1.7e-11 Length: 767
Pred. No.: 263.50 Matches: 125
Score: 40.45% Conservative: 93
Percent Similarity:

```

Best Local Similarity: 23.19%      Mismatches: 206
Query Match: 6.40%               Indels: 115
DB: 10                           Gaps: 24

US-09-502-945-4 (1-2236) x US-09-919-497-59 (1-767)

QY 340 TCCAGAGAGCTGAAGAGGTGCTCGACCGCTCTCCACCCCAAGCGCTCGCGCTGAGT 399
D 198 AlslulysValMetGluilleuLysLeuLys---GlyProlysGlyLeuGlyPheSer 216
QY 400 GTCGCTGGTGGC-----CTGGAGCTTGGCTGGCTCTCACTCCAC 444
D 217 IleAlaGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleYrValThrLys 236
QY 445 CTCATCAAGCGGTGAGGAGACAGCGTCGGG---CTCAGTAGGGGAGGATCGTC 501
D 237 IleleGluGlyAlaAlaHisLysAspGlyArgLeuGlnleGlyAspLysIleLeu 256
QY 502 CGGATCAATGGATATTCCTCTCTCTGTCACCATGAGGAGTCATCACTCATCGA 561
D 257 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaAlaLeuLys 276
QY 562 ---ACCAAGAAACTGTGTCATCAAGTG-----AGACATCGCGCTGATCCCC 609
D 277 AsnThrTyraSpValValTyLeuLysValAlaLysProSerAsnAlaTyLeuSerAsp 296
QY 610 GTGAAAGCTCTCTGATGAGCCCTCCTACTTGG---CAGTATGTGATCAGTTGTGTCG 666
D 297 SerTyraAlaProAspIleThrThrSerTyraSerGlnHisAspAsnGluIleSer 316
QY 667 GAATCTGGGGCGTGGAGCAGCCCTGGC----- 696
D 317 HisSer-----SerTyraLeuGlyThrAspTyraProThrAlaMetThrProThr 332
QY 697 -----TCCCTCGAATCGGAA-----AACAG 720
D 333 SerProArgArgTyraSerProValAlaLysAspLeuLeuGlyLysValProArg 352
QY 721 GAGAAGAGTCTATCAGCCCTGGTAGGCTCCCGAGGCTGGTGGCAGCAATTCAGAC 780
D 353 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 372
QY 781 GGCCCCATCCAGAGCGTGCATCTTTATCAGCCATGTG-----AAACCTGCTCC 831
D 373 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 391
QY 832 CTGCTCTGCTGAGTGGGATGGAGATGAGGACAGATGTCGAATGCTCAATGCGCTCGAC 891
D 392 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 409
QY 892 TTCTCTAACCTGGATCACAAGGAGGCTGTAATGTCTCTGCTGAAATAGCCGACGCTGACC 951
D 410 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 428
QY 952 ATCTCCATGTAGCTCAGCTGCCCGGAGCTGTTATCATCAGACCGGCGGCTGGCA 1011
D 429 ValThrIleIleAlaGln-----TyraLeuGlyLeuTyraSerArgPhe 443
QY 1012 GAGCGCGGCGAGCTGAGCTGAGCGGCGAGGAGCTCTCATCAGAGCGGCTGGCGATG 1071
D 444 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 461
QY 1072 GAGTCCACAAGATCTCCAGGAGCAGCAGGAG-----ATGGAG 1110
D 462 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyraIleArgAlaLeuPheAsp 481
QY 1111 CGGCAAGAGCAAAAGAAATTCGCCAGAGGAGCAGAGCAAAATGAGATACCGGAAG 1170
D 482 TyraSpLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 501
QY 1171 GAGATGGAACAGATGTACAGGAGGAGGAGAGAGTGT----- 1206
D 502 ValLeuHisValIleAspAlaSerAspGluGluTrpTrpGluAlaArgValHisSer 521

```

```

QY 1207 -----AAGAAGCAA 1215
D 522 AspSerGluThrAspAspIleGlyPheIleProSerLysArgValGluArgGlu 541
QY 1216 TGGGAA-----GAAGACTGGGCTCA-----AAGGAA 1242
D 542 TrpSerArgLeuLysAlaLysAspTrpGlySerSerGlySerGlnGlyArgGluAsp 561
QY 1243 CAGCTACTCTTCCTTAAACCATCACT---GCTGAGGTACAC-----CCAGTACCC 1290
D 562 SerValLeuSerTyraGluThrValThrGlnMetGluValHisTyraAlaArgProIleIle 581
QY 1291 CTTCCCAAGCCCAAGTATGATCAGGAGTGGAACTCGAGCTCGAGCCCGCAGATGCTG 1350
D 582 IleLeuGlyProThrLysAspArgAlaAsnAspAspLeuLeuSerGluPheProAspLys 601
QY 1351 GATGAGGCGACGAGGAGGAGGAGGAGGAGGATTTCCGGAAATATGAGGAGGCTTTGAC 1410
D 602 PheGlySerCysValProHisThrThrArgProLysArgGluTyraGluIleAspGlyArg 621
QY 1411 CCTACTCTATCTTCCACCCAGACGATCATGCGGAAGGATGTCGGCTCCTACGATC 1470
D 622 AspTyraHisPheValSerSerArgGluLysMetGluLysAspIleGlnAlaHisLysPhe 641
QY 1471 AAGAAGGAGGATCCTTAGACCTGCGCTCGGAAAGCGGTGTGGACTCCCCCATTT----- 1524
D 642 IleGluAlaGlyGlnTyraAsnSerHisLeuTyraGlyThrSerValGlnSerValArgGlu 661
QY 1525 -----GGGAG-----GTGCTGTTCTGCTGTGTATGAGCGG 1557
D 662 ValAlaGluGlnGlyHisCysIleLeuAspValSerAlaAsnAlaValArgArgLeu 681
QY 1558 GGAGCTGCTGAGCGCATGCTGGCATTTGTG-----AAAGGGACGAGATC 1602
D 682 GlnAlaAlaHisLeuHisProIleAlaIlePheIleArgProArgSerLeuGluAsnVal 701
QY 1603 ATGGCAATCAACGCGCAAGATTTGTACAGACTACACCTGGCTGGAGCTGAGCTGCC 1659
D 702 LeuGluIleAsnLysArgIleThrGluGluAlaArgLysAlaPheAspArgAla 720

RESULT 12
US-09-998-425-3
; Sequence 3, Application US/09998425
; Publication No. US20030008346A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavitlian, Sean V.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: MMSC1 - An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC1 Gene
; CURRENT APPLICATION NUMBER: US/09/998,425
; CURRENT FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/233,086
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/071,861
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-998-425-3

Alignment Scores:
Pred. No.: 4,99e-11 Length: 1881
Score: 259.50 Matches: 149
Percent Similarity: 37.03% Conservative: 115
Best Local Similarity: 20.90% Mismatches: 232
Query Match: 6.30% Indels: 217
DB: 9 Gaps: 31

```


Search completed: March 21, 2003, 13:57:52
Job time : 53.9422 secs


```

Db 174 lleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 193
QY 445 CTCACAAAGCGGTGACGACAGCTCGG---CTCCAGGTAGGGACGACATCGTC 501
Db 194 lleIleGluGlyGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 213
QY 502 CGGATCAATGATATTCATCTCTCTGTCACCATGAGGAGGTCAATCAACCTCATTCGA 561
Db 214 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaLeuLys 233
QY 562 ---ACCAGAAAAGTGTCTCCATCAAAAGT-----AGACACATCGGCTGATCCCC 609
Db 234 AsnThrTyrAspValValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 253
QY 610 GTGAAAAGCTCTCCTGATGAGCCCTCACTGG---CAGTATGGATCAGTGTGTGTCG 666
Db 254 SerTyrAlaProProAspIleThrSerTyrSerGlnHisLeuAspAsnGluIleSer 273
QY 667 GAATCTGGGGCGTGGAGGCGCTGGCC-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 289
Db 274 HisSer-----TCCCTGGAATCGGAA-----AACAAAG 720
QY 721 GAGAAAGGCTTTCATCAGCTGCTAGCTCCCGAGGCTGGCTGAGCATTTCCAGC 780
Db 310 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 329
QY 781 GCCCCATCCAGAACCTGGCATTTTATCAGCCATGTG-----AAACCTGGCTCC 831
Db 330 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 348
QY 832 CTGCTGCTGAGTGGGATGGAGTGGGACACAGATGTCGAAGTCAATGGCGCTCAGC 891
Db 349 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 366
QY 892 TTCTCTCACTGCATCACAGAGGCTGTAAATGTGCTGAAATAGCGGACGCTGACC 951
Db 367 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyIleThr 385
QY 952 ATCTCCATGTAGTCACTGAGCTGGCGGGAGCTGTTTCATCAGACAGCGGGAGCGGTGCA 1011
Db 386 ValThrIleIleAlaGln-----TyrLysProGluLutyrSerArgPhe 400
QY 1012 GAGCGCGGCGCTGAGCTGAGCGGCGGAGGCTTCTCATCCAGACGGCGTGGCGATG 1071
Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
QY 1072 GAGTCCAAACAAGATCTCCAGGACGACGAGGAG-----ATGGAG 1110
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
QY 1111 CGGCAAAAGGAGAAAGAAATATCCAGAGGCGGACGAGGAGAAATCAGAGATACCGGAG 1170
Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 458
QY 1171 GAGATGGAACATGTAGAGGAGGAGGAGAGT-----1206
Db 459 ValLeuHisValIleAspAlaGlyAspGluGluTyrTrpGlnAlaArgValHisSer 478
QY 1207 -----AAGAGGAA 1215
Db 479 AspSerGluThrAspAspIleGlyPheIleProSerLysArgValGluArgArgGlu 498
QY 1216 TGGGAAGAACTGGGGCTCA-----AAGAACAGCTACTCTTG 1254
Db 499 TrpSerArgLeuTrpGlySerSerGlySerGlnGlyArgGluAspSerValLeuSer 518
QY 1255 CCTAAACCATCACT---CCTGAGGTACAC-----CCAGTACCCCTTGGCAGGCA 1302
Db 519 TyrGluThrValThrGlnMetGluValHisTyrAlaArgProIleIleLeuGlyPro 538

```

```

QY 1303 AAGTATGATCAGGAGTGGAACTGAGCTCGAGCCGACAGATGATGATGAGGACG 1362
Db 539 ThrLysAspArgAlaAsnAspAspLeuLeuSerProSerProAspLysPheGlySerCys 558
QY 1363 GAGGACGAGGAGCAGGAGTATTCGGAATATGAGGAAGCTTTGACCCCTACTCTATG 1422
Db 559 ValProHisThrThrArgProLysArgGluTyrGluIleAspGlyArgAspTyrHisPhe 578
QY 1423 TTCACCCAGCAGCAGATCATGGGAGGAGTCTCGGCTCCTACGCATCAGAGAGGGA 1482
Db 579 ValSerSerArgGluLysMetGluLysAspIleGlnAlaHisLysPheIleGluAlaGly 598
QY 1483 TCCTTAGACCTGGCCCTGGAAGCGGTGGACTCCCCCAT-----1524
Db 599 GlnTyrAsnSerHisLeuTyrGlyThrSerValGlnSerValArgGluValAlaGluGln 618
QY 1525 GGAAG-----GTGGTGTCTTCTGTATGATGAGCGGGAGCTGTGTGAG 1569
Db 619 GlyLysHisCysIleLeuAspValSerAlaAsnAlaValArgArgLeuGlnAlaAlaHis 638
QY 1570 CGGCATGGTGCATTGTG-----AAAGGGACGAGATCATGGCAATCAAC 1614
Db 639 LeuHisProIleAlaIlePheIleArgProArgSerLeuGluAsnValLeuGluIleAsn 658
QY 1615 GGCAGATTTGTACAGACTACACCTGGCTGAGCTGACGCTGCC 1659
Db 659 LysArgIleThrGluGluGlnAlaArgLysAlaPheAspArgAla 673

RESULT 2
T46612
multi PDZ domain protein 1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 18-Feb-2000 #sequence_revision 18-feb-2000 #text_change 21-Jul-2000
C:Accession: T46612
R:Ullmer, C.; Schmuck, K.; Figge, A.; Lubbert, H.
FEBS Lett. 424, 63-68, 1998
A:Title: Cloning and characterization of MUPP1, a novel PDZ domain protein.
A:Reference number: 223104; MUID:98196865; PMID:9537516
A:Accession: T46612
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-2054 <ULL>
A:Cross-references: EMBL:AJ001320; NID:g2959978; PIDN:CAA04681.1; PID:g2959979
A:Experimental source: brain
C:Genetics:
A:Gene: MUPP1

Alignment Scores:
Pred. No.: 4.63e-09 Length: 2054
Score: 269.00 Matches: 120
Percent Similarity: 36.64% Conservative: 72
Best Local Similarity: 22.90% Mismatches: 160
Query Match: 6.53% Indels: 172
DB: 2 Gaps: 17

US-09-502-945-4 (1-2236) x T46612 (1-2054)
QY 157 GCAGAGAGGAGTATCTCTATGATGCTGGGAATGTACACACAGACCATGGAGCTGCC 216
Db 1672 AlaThrHisAspGluAlaIleAsnValLeuArgGlnThrProGlnArgVal-----1688
QY 217 GTGCTCTGGGACAGCTGAAGCTGTCATCAATGAACCCAGCCGCTCTGCTCTGTTGAT 276
Db 1689 -----ArgLeuThrLeuTyr---1693
QY 277 GCCATTCCGCGCTGATCCCACTGAAGCACCAGCTGGAATATGATCAGTACCCCGG 336
Db 1694 -----ArgAspGluAlaProTyrLysGluAspValCysAspThrPheThr-----1709
QY 337 CGCTCCAGGAAGCTGAGGAGGCTGCTGAGCGTCTGCACCCCGAGGCGCTCGGCTG 396
Db 1710 -----ValGluLeuGlnLysArgProGlyLysGlyLeuGlyLeu 1722

```

QY 397 AGTGTCCGTGGTCCCTGGAGTTGGCTGTGGCTCTTTCATCTCCACACTCATCAAGGC 456
 Db 1723 SerIleValGlyLysArgAsn---AspThrGlyValPheValSerAspIleValLysGly 1741
 QY 457 GGTACGGCAGACAGCGCTCGGG---CTCCAGGTAGGGAGGAGATCTCCGGATCAATGA 513
 Db 1742 GlyIleAlaAspAlaAspGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGly 1761
 QY 514 TATTCATCTCTCTGTACCATGAGGAGGTCATCAACTTCATTCGA---ACCAAGAAA 570
 Db 1762 GluAspValArgAsnAlaThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGly 1781
 QY 571 ACTGTCTCCATCAAGTGAAGACATCGCCCTGATCCCTCGTGAAGAGCTCTCTGATGAG 630
 Db 1782 ThrValThrLeuGluValGlyArgIleLysAlaAlaPheHisSerGluArgArgPro 1801
 QY 631 CCCCTCACTGGCAGTATGTGGATCAGTTGTGTGCGAATCTGGGGCGTGGAGGCAGC 690
 Db 1802 SerGlnSerSerGlnValSerGluSerSerLeuSerSerPheSerLeuProArgSerGly 1821
 QY 691 CTGGGCTCCCTGGAAATCGGAAACAAGAGAGAAG--- 729
 Db 1822 IleHisThrSerGluSerSerGluSerSerAlaLysLysAsnAlaLeuAlaSerGluIle 1841
 QY 730 -----GTCTTCATCAGCCCTGGTAGGCTCCGAGGCTGGCTGCGAGCAT 774
 Db 1842 GlnGlyLeuArgThrValGluIleLysLysGlyProAlaAspAlaLeuSerIle 1861
 QY 775 TCCAGCGGC-----CCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAACCT 825
 Db 1862 AlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetMethHisPro 1881
 QY 826 GGCCTCCCTCTCTCAGGTG---GGATTGGAGATAGGGACAGATGTCCGAATCAAT 882
 Db 1882 AsnGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleCys 1901
 QY 883 GCGCTCGACTTCTTAACCTGGATCACAGGAGGCTGTAATGTCTGAAATAGCGGC 942
 Db 1902 GlyThrSerThrAspGlyMetThrIsthrGlnAlaValAsnLeuMetLysAsnAla--- 1920
 QY 943 AGCCTGACCATCTCCATCTAGCTGAGCTGGCGGGAGCTGTTTCATGACAGCCGGAG 1002
 Db 1921 SerGlySerIleGluValGlnValAlaGlyGlyAspVal----- 1934
 QY 1003 CGGCTGGCAGAGCGCGGCGAGCTGAGCTGACGGCAGGAGCTTCTCATGACAGAGCGG 1062
 Db 1935 SerValValThrGlyHisGlnGlnGluLeuAlaAsnProCysLeuAlaPhe----- 1951
 QY 1063 CTGGCGATGGATCCCAACAGATCTCCAGGAGCAGCAGGAGATCGAGCGGCAAGGAGA 1122
 Db 1952 ThrGlyLeuThrSerSerThrIle----- 1959
 QY 1123 AAAGAAATGCCAGAGGACGACAGAGGAAAATGAGATACCGGAAGAGATGAACAG 1182
 Db 1959 ----- 1959
 QY 1183 ATTGTAGAGGAGAGAGAAGTTTAAGAAGCAATGGGAAGAGACTGGGGCTCAAGGAA 1242
 Db 1960 -----PheProAspAspLeuGlyProProGln 1968
 QY 1243 CAGCTACTCTTGCCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCA 1302
 Db 1969 Ser-----LysThrIleThr----- 1973
 QY 1303 AAGTATGATCAGGAGTGAACCTGAGCTGAGCCCGCAGATGACCTGGATGGAGGCAGC 1362
 Db 1974 -----LeuAspArgGlyPro 1978
 QY 1363 GAGGACGAGGAGACGAGGATTTCCGGAATATGAGGAAGGCTTGACCCCTACTCTATG 1422
 Db 1979 Asp----- 1979

QY 1423 TTCACCCAGAGCAGATCATGGGAAGGATGTCCGGCTCTACGCATCAAGAGAGGGA 1482
 Db 1979 ----- 1979
 QY 1483 TCCTTAGACCTGGCCTGGAAGGGGTGTGGACTCCCCCATTTGG- 1536
 Db 1980 GlyLeuGlyPheSerIleValGlyGlyTyrosGlySerProHisGlyAspLeuProIleTyr 1999
 QY 1537 GTTCTGTGTATGACGGGGAGCTGCTGAGCGGATGCTGGCATTTGTGAAAGGGGAC 1596
 Db 2000 ValLysThrValPheAlaLysGlyAlaAlaGluAspGlyArgLeuLysArgGlyAsp 2019
 QY 1597 GAGATCATGGCAATCAAGCGCAAGATGTGTGACAGACTACACCTGGCTGAGGCTGACGCT 1656
 Db 2020 GlnIleIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaValAla 2039
 QY 1657 GCCCTGCAGAG 1668
 Db 2040 IleLeuLysArg 2043
 RESULT 3
 T09599
 postsynaptic density protein 95 - human
 C:Species: Homo sapiens (man)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
 C:Accession: T09599
 R:Stathakis, D.G.; Hoover, K.H.; You, Z.; Bryant, P.J.
 submitted to the EMBL Data Library, July 1998
 A:Reference number: Z16761
 A:Accession: T09599
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-767 <STA>
 A:Cross-references: EMBL:U83192; NID:g3318652; PID:g3318653
 A:Experimental source: mammary
 C:Genetics:
 A:Gene: PSD95
 C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
 F:208-286/Domain: GLGF domain homology <GLG>
 F:478-536/Domain: SH3 homology <SH3>
 F:578-755/Domain: guanylate kinase homology <GKI>

Alignment Scores:
 Pred. No.: 7.36e-09 Length: 767
 Score: 263.50 Matches: 125
 Percent Similarity: 40.45% Conservative: 93
 Best Local Similarity: 23.19% Mismatches: 206
 Query Match: 6.40% Indels: 115
 DB: 2 Gaps: 24

US-09-502-945-4 (1-2236) x T09599 (1-767)

QY 340 TCCAGGAGCTGAAGGAGGTGCGTCTGGACGCTCTGCACCCGAGGCGCTCGGCTCAGT 399
 Db 198 AlaGluLysValMetGluIleLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 216
 QY 400 GTGCGTGGTGGC-----CTGGAGTTTGGCTGTGGCTCTTCATCTCCAC 444
 Db 217 IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 236
 QY 445 CTCATCAAGCGGTGACGAGCAGAGCGTGGG---CTCAGGTAGGGGAGGAGATCGTC 501
 Db 237 IleIleGluGlyGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 256
 QY 502 CGGATCAATGGATTTCCATCTCTCTGACCCATGAGGAGGTATCAACCTCATTCGA 561
 Db 257 AlaValAsnSerValGlyLeuGluAspValMethHisGluAspAlaValAlaLeuLys 276
 QY 562 ---ACCAAGAAACTGTGTCTCCATCAAGTG-----AGACACATCGCCTGATCCCC 609
 Db 277 AsnThrThrAspValValThrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 296
 QY 610 GTGAAAAGCTCTCTGATGAGCCCTCCTCTGG---CAGTATGTGGATCAGTTTGTGTGCG 666

```
Db 297 SerTyrAlaProAspIleThrThrSerTyrSerGlnHisLeuAspAsnGluIleSer 316
QY 667 GAATCTGGGGCGCTGCAGCGACCTGGGC----- 696
Db 317 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 332
QY 697 -----TCCCTCGAATCGGAA-----AACAG 720
Db 333 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluAlaAspIleProArg 352
QY 721 GAGAGAAGGTCTTCATCAGCTGTAGCTCCGAGGCGCTGGCTGCAGCATTTCCAGC 780
Db 353 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 372
QY 781 GGCCTCATCCAGAAAGCTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
Db 373 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 391
QY 832 CTCTCTGCTGAGTGGATTGAGATGAGGGACCATGTTGCAAGTCAATGCGGTGCGAC 891
Db 392 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 409
QY 892 TTCTCTAACTGTCATCACAGGAGGTGTAATGTCTGAAATAAGCCGACCTGACC 951
Db 410 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 428
QY 952 ATCTCCATTGTAGCTGACGTGGCCGGGAGCTGTTTCATGACAGACCGGGAGCGCTGGCA 1011
Db 429 ValThrIleIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 443
QY 1012 GAGCGCGGCGAGCTGAGCTGACGCGGAGGAGCTTCTCATCAGAGAAGCGGTGGCGATG 1071
Db 444 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 461
QY 1072 GAGTCCAAACAAGATCTCCAGGAGCAGCAGAG-----ATGGAG 1110
Db 462 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 481
QY 1111 CGCAAGAGGAGAAAGAAATTCGCCAAGAGCAGCAGAGGAAATAGAGATACCGGAAG 1170
Db 482 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 501
QY 1171 GAGATGGAACAGATTGTAGAGGAGGAGAGACTTT----- 1206
Db 502 ValLeuHisValIleAspAlaSerAspGluGluTrpTrpGlnAlaArgValHisSer 521
QY 1207 ----- 1215
Db 522 AspSerGluThrAspAspIleGlyPheIleProSerLysArgArgValGluArgGlu 541
QY 1216 TGGGAA-----GAAGACTGGGGCTCA-----AAGGAA 1242
Db 542 TrpSerArgLeuLysAlaLysAspTrpGlySerSerSerGlySerGlnGlyArgGluAsp 561
QY 1243 CAGCTACTCTTGCTTAAACCATCACT---GCTGAGGTACAC-----CCAGTACCC 1290
Db 562 SerValLeuSerTyrGluThrValThrGlnMetGluValHisTyrAlaArgProIleIle 581
QY 1291 CTTCGCAAGCCAAAGATGATGATGAGGAGTGGAACTGCTGAGCTCAGCCGCGAGATGACCTG 1350
Db 582 IleLeuGlyProThrLysArgAlaAsnAspLeuLeuSerGluPheProAspLys 601
QY 1351 GATGGAGGCACGGAGGAGCAGGAGCAGGATTTCCGAAATATGAGGAAGCTTTGAC 1410
Db 602 PheGlySerCysValProHisThrThrArgProLysArgGluTyrGluIleAspGlyArg 621
QY 1411 CCTACTCTATGTTACCCAGACAGATCATCGGGAAGATCTCCGGTCTCTACGCATC 1470
Db 622 AspTyrHisPheValSerSerArgGluLysMetGluLysAspIleGlnAlaHisLysPhe 641
QY 1471 AAGAGGAGGATCCTTAGACCTGGCCCTGGGAAGGGGTGTGGACTCCCCCATTT----- 1524
```

```
Db 642 IleGluAlaGlyGlnTyrAsnSerHisLeuTyrGlyThrSerValGlnSerValArgGlu 661
QY 1525 -----GGGAAG-----CTGCTCGCTTTCTGCTGTGTATGACGG 1557
Db 662 ValAlaGluGlnGlyLysHisCysIleLeuAspValSerAlaAsnAlaValArgArgLeu 681
QY 1558 GGAGCTGCTAGCGCGCATGTGGCATTTGTG-----AAAGGGACGAGATC 1602
Db 682 GlnAlaAlaHisLeuHisProIleAlaIlePheIleArgProArgSerLeuGluAsnVal 701
QY 1603 ATGGCAATCAACGCAAGATTGTGACAGCTACACCTGGCTGAGCGCTGCGCTGCC 1659
Db 702 LeuGluIleAsnLysArgIleThrGluGluGlnAlaArgLysAlaPheAspArgAla 720
RESULT 4
JH0800
postsynaptic density protein PSD-95 - rat
N:Alternate names: brain specific PSD-95 protein; discs-large tumor suppressor protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C:Accession: JH0800; S26407
R:Cho, K.O.; Hunt, C.A.; Kennedy, M.B.
Neuron 9, 929-942, 1992
A:Title: The rat brain postsynaptic density fraction contains a homolog of the drosophila discs-large tumor suppressor protein
A:Reference number: JH0800; MUID:93040233; PMID:1419001
A:Accession: JH0800
A:Molecule type: mRNA
A:Residues: 1-724 <CHO>
A:Cross-references: GB:M96853; NID:g206454; PIDN:AAA41971.1; PID:g206455
A:Experimental source: brain
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
C:70-148/Domain: GLGF domain homology <GLG1>
F:163-243/Domain: GLGF domain homology <GLG2>
F:435-493/Domain: SH3 homology <SH3>
F:535-712/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 9,67e-09 Length: 724
Score: 261.50 Matches: 125
Percent Similarity: 40.45% Conservative: 93
Best Local Similarity: 23.19% Mismatches: 206
Query Match: 6.35% Indels: 115
DB: 2 Gaps: 24
US-09-502-945-4 (1-2236) x JH0800 (1-724)
QY 340 TCCAGAAAGCTGAAGAGGTGCTGCTGACCGCTGTGACCCGAGGCGCTGGCTGAGT 399
Db 155 AlaGluLysValMetGluIleLysLeuIleLys---GlyProLysGlyLeuGlyPheSer 173
QY 400 GTGCTGTGGTGC-----CTGAGTTTGGCTGTGGGCTCTTCATCTCCAC 444
Db 174 IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 193
QY 445 CTCATCAAGCGGTGAGCAGCAGCAGCTCGG---CTCCAGTAGGAGGAGATCTGTC 501
Db 194 IleIleGluGlyGlyAlaHisLysAspGlyArgLeuGlnIleGlyLysIleLeu 213
QY 502 CGGATCAATGGATATTCATCTCCCTGTACCCATGAGGAGGTCAACCTCATTCGA 561
Db 214 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaLeuLys 233
QY 562 ---ACCAAGAAACTGTCTCCATCAAGTG-----AGACATCGGCTGTATCCCC 609
Db 234 AsnThrTyrAspValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 253
QY 610 GTGAAAGCTCTCTGATGAGCCCTCACTGG---CAGTATGTCAGTCAGTTTGTGTCG 666
Db 254 SerTyrAlaProProAspIleThrThrSerTyrSerGlnHisLeuAspAsnGluIleSer 273
QY 667 GAATCTGGGGCGTCCGAGGAGCGCTGGC----- 696
Db 274 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 289
```

```
Qy 697 -----TCCCTCGAATCGGAA-----AACAG 720
Db 290 SerProArgArgTyrSerProValAlaLysAspLeuLeuGluGluAspIleProArg 309
Qy 721 GAGAAAGAGGTCTTCATCAGCGCTGTAGCTCCCGAGGCTGGCTGCGAGCATTTCCAGC 780
Db 310 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 329
Qy 781 GCCCCATCCAGACCTGGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
Db 330 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 348
Qy 832 CTGTCTGTCTGAGTGGGATGGAGATAGGGACCGAGATGTCGAAGTCAATGGCTGCAC 891
Db 349 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 366
Qy 892 TTCTCTAACCTGGATCAACAGGAGGCTGTAAATGTGCTGAAATAGCGGAGCGCTGACC 951
Db 367 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 385
Qy 952 ATCTCCATGTAGCTGAGCTGCGCGGAGGCTGTTTCATCAGACCGCGGAGCGCTGGCA 1011
Db 386 ValThrIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
Qy 1012 GAGGCGCGCAGCTGAGCTGCGCGGAGGAGCTTCATGCGAGAACGGGTGGCGATG 1071
Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
Qy 1072 GAGTCCAAACAGATCTCCAGGAGCAGCAGGAG-----ATGCAG 1110
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
Qy 1111 CGGAAAGAGGAAAGAAATGCCCAGAGGAGCAGGAGGAGAAATGAGAGATACCGGAAG 1170
Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 458
Qy 1171 GAGATGGAACAGATGTAGAGGAGGAGAGAGAGTTT-----1206
Db 459 ValLeuHisValIleAspAlaGlyAspGluGluIleTyrPheGlnAlaArgValHisSer 478
Qy 1207 -----AAGAACAA 1215
Db 479 AspSerGluThrAspAspIleGlyPheIleProSerLysArgArgValGluArgArgGlu 498
Qy 1216 TGGNA-----GAGACTGGGCTCA-----AAGAA 1242
Db 499 TrpSerArgLeuLysAlaLysAspTrpGlySerSerSerGlySerGlnGlyArgGluAsp 518
Qy 1243 CAGTACTCTTCCTAAACCATCACT---GCTGAGGTACAC-----CCAGTACCC 1290
Db 519 SerValLeuSerTyrGluThrValThrGlnMetGluValHisTyrAlaArgProIleIle 538
Qy 1291 CTTCCAGACCAAGATATGATCAGGAGTGGAACTGAGCTGAGCCCGCAGATGACCTG 1350
Db 539 IleLeuGlyProThrLysAspArgAlaAsnAspAspLeuLeuSerGluPheProAspLys 558
Qy 1351 GATGAGCAGCAGGAGCAGGAGCAGGAGGATTTCCGGAATATGAGGAAGGCTTTGAC 1410
Db 559 PheGlySerCysValProHisThrThrArgProLysArgGluTyrGluIleAspGlyArg 578
Qy 1411 CCCTACTCTATGTTACCCAGCAGATCATGCGGAGGAGGATGTCGGCTCCTACGCATC 1470
Db 579 AspTyrHisPheValSerSerArgGluLysMetGluLysAspIleGlnAlaHisLysPhe 598
Qy 1471 AAGAGGAGGATCTTAGACTGCGCTGGAAGCGGTGTGGACTCCCGCAT-----1524
Db 599 IleGluAlaGlyGlnTyrAsnSerHisLeuTyrGlyThrSerValGlnSerValArgGlu 618
Qy 1525 -----GGGAAG-----GTGGTCGTTTCTGCTGTATGAGCGG 1557
Db 619 ValAlaGluGlnGlyLysHisCysIleLeuAspValSerAlaAsnAlaValArgArgLeu 638
```

```
Qy 1558 GGAGCTCTGAGCGGCATGTGGTCATTTGTG-----AAAGGGACGACATC 1602
Db 639 GlnAlaAlaHisLeuHisProIleAlaIlePheIleArgProArgSerLeuGluAsnVal 658
Qy 1603 ATGGCAATCAACGGCAAGATTTGTGACAGACTACACCTGGCTGAGGCTGACGCTGCC 1659
Db 659 LeuGluIleAsnLysArgIleThrGluGluGlnAlaArgLysAlaPheAspArgAla 677
RESULT 5
A46431
tight junction-associated protein ZO-1 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A46431
R:Itoh, M.; Nagatuchi, A.; Yonemura, S.; Kitani-Yasuda, T.; Tsukita, S.
J. Cell Biol. 121, 491-502, 1993
A:Title: The 220-kD protein colocalizing with cadherins in non-epithelial cells is id
py.
A:Reference number: A46431; MUID:93252986; PMID:8486731
A:Accession: A46431
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1745 <ITO>
A:Cross-references: GB:D14340; NID:g303709; PIDN:BAA03274.1; PID:d1003784; PID:g30371
A:Experimental source: GB:F9 cells
A:Note: sequence extracted from NCBI backbone (NCBIN:131200, NCBIP:131201)
C:Superfamily: guanylate kinase homology; GIGF domain homology
F:27-106/Domain: GIGF domain homology <GIG1>
F:428-498/Domain: GIGF domain homology <GIG3>
F:645-794/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 1.31e-08 Length: 1745
Score: 261.50 Matches: 143
Percent Similarity: 36.73% Conservative: 106
Best Local Similarity: 21.09% Mismatches: 222
Query Match: 6.35% Indels: 207
DB: 2 Gaps: 30
US-09-502-945-4 (1-2236) x A46431 (1-1745)
Qy 358 GTGGCTGTGGACCTCTGCACCCGAGGCTCGCGCTGAGTGTGGTGGCTGGAG 417
Db 24 ValThrLeuHisArgAlaProGlyPheGlyIleAlaIleSerGlyGlyArgAsp 43
Qy 418 -----TTTGCTGTGGG-----CTCTTCATCTCCCACTCAACAGGCGGT 459
Db 44 AsnProHisPheGlnSerGlyGluThrSerIleValIleSerAspValLeuLysGlyGly 63
Qy 460 CAGGCACAGCAGCTGGGCTCCAGGTAGGGACAGATCGTCCGGATCAATGGATATCC 519
Db 64 ProAlaGluGly---GlnLeuGlnLysAsnAspArgValAlaMetValAsnGlyValSer 82
Qy 520 ATCTCTCTCTTACCCATGAGGAGGTCTCAAC-----552
Db 83 MetAspAsnValGluHisAlaPheAlaValGlnGlnLeuArgLysSerGlyLysAsnAla 102
Qy 553 -----CTCATTTCAACCAAGAACTGTGTCCATCAAGTGAGACATCGGCTGATC 606
Db 103 LysIleThrIleArgArgLysLysValGlnIleProValSerHis-----118
Qy 607 CCCGTGAAAGCTCTCCTGAT---GAGCCCTCACTTTGGCAGTATGGGATACATTTGTG 663
Db 119 -----ProAspProGluProValSerAspAsnGluAspAspSerTyrAsp 133
Qy 664 TCGAATCTGGGGCGTGGAGGAGCAGCTGGCTCCCTCGAATCGGAAACAAAGAG 723
Db 134 GluGluValHisAspProArgAlaGlyArgGlyAlaLeuAlaAsnArgArgSerGlyLys 153
Qy 723 -----723
Db 154 SerTrpAlaArgAspArgSerAlaSerArgGluArgSerLeuSerProArgSerAspArg 173
```

```
QY 724 -----AAGAGGTCTTCATCAGCCTGGTAGGCTCCGA 756
Db 174 ArgSerValAlaSerSerGlnProAlaLysProThrLysValThrLeuValLysSerArg 193
QY 757 -----GGCTTGGTGCAGCATTCACGGGCCCATCCAGAACGCTGGC 801
Db 194 LysAsnGluGluTyrGlyLeu-----ArgProAla 203
QY 802 -----ATCTTTATCAGCCATCGAAACCTGGCTCCCTGCTGCTGAGGTGGGA---TTG 852
Db 204 SerHisIlePheValLysGluIleSerGlnAspSerLeuAlaAlaAspGlyAspIle 223
QY 853 GAGATAGGGGACCGATGTGGAAGTCAATGGCGTCGACTTCTCTAACCCTGGATCACAG 912
Db 224 GlnGluGlyAspValValLeuLysIleAsnGlyThrValThrGluAsnMetSerLeuThr 243
QY 913 GAGGCTGTAAATGTCTGAAATAATAGCGCAGCCTGACCATCTCCATGTAGTGCAGCT 972
Db 244 AspAlaLysThrLeuIleGluArgSerLysGlyLysLeuLysMetValValGlnArgAsp 263
QY 973 GGCCGGGAGCTGTTTCATG-----990
Db 264 GluArgAlaThrLeuLeuAsnValProAspLeuSerAspSerIleHisSerAlaAsnAla 283
QY 991 ACAGACCGGAGCGGCTGCGACAGCGCGGCGAG-----CGTGAG 1029
Db 284 SerGluArgAspAspIleSerGluIleGlnSerLeuAlaSerAspHisSerGlyArgSer 303
QY 1030 CTGACCGCGCAGGACTTCTCATGCAGAGCGGCTG-----1065
Db 304 HisAspArgProArgArgSerGlnSerArgSerProAspGlnArgSerGluProSer 323
QY 1066 -----CGCATGGAGTCCAAAGATCTCCAGGAGCAGCAGGAG 1104
Db 324 AspHisSerThrGlnSerProGlnProSerAsnGlySerLeuArgSerArgGluGlu 343
QY 1105 ATGGAGCGCAAGGAGAAAGAAATTGCCAGAGGCGCAGAGAGAAATGAGAGATAC 1164
Db 344 GluArgMetSerLysProGlyAlaIleSerThrProValLysHisValAspAspHisPro 363
QY 1165 CGGAGGAGATGCACAGATT---GTAGAGGAGGAGAGCAAGTAAAGAACCAATGGGAA 1221
Db 364 ProLysAlaValGluGluValThrValGluLysAsnGluLys-----377
QY 1222 GAAGACTGGGGCTCAAAGGAACAGCTACTCTTGCCTAAACCATCATCTGCTGAGTACAC 1281
Db 378 -----GlnThrPro 380
QY 1282 CCAGTACCCCTTCGAAGCAAGATATGATCAGGAGTGGAACTGAGCTCGAG---CCC 1338
Db 381 ThrLeuProGluProLysProValTyrAlaGlnValGlyGlnProAspValAspLeuPro 400
QY 1339 GCAGATGACCTGGAGGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1398
Db 401 ValSerProSerAspGlyAlaLeuProAsnSerAlaHisGluAsp-----415
QY 1399 GAAGGCTTTGACCCCTACTCTATGTTTACCCCGCAGACGATCATGGGGAAGGATGCGGG 1458
Db 416 -----GlyIleLeuArgProSerMetLys 423
QY 1459 CTCCTACGCATCAAGAGGAGGATCCTTTAGACCTGGCCCTGGAAGCGGCTGGACATCC 1518
Db 424 LeuValLysPheArgLysGlyAspSerValGlyLeuArgLeuAlaGlyLysAsnAsp---442
QY 1519 CCCATTGGGAAGGTGCTCTTCTGCTGTATGACGGGGGAGCTGCTGAGCGGATGGT 1578
Db 443 ---ValGly---IlePheValAlaGlyValLeuGluAspSerProAlaAlaLysGlu---459
QY 1579 GGCATTGTGAAGGAGCAGAGATCATGGCAATCAACGGCAAGATTCTGACAGACTACACC 1638
Db 460 GlyLeuGluGlyAspGlnIleLeuArgValAsnAsnValAspPheThrAsnIleIle 479
QY 1639 CTGGCTGAGGCTGACGCTGCCCTGCAGAGGCGCTGGAATCATGCGGCGGAGCTGGATCGAC 1698
```

```
Db 480 ArgGluGluAlaValLeuPheLeu-----LeuAsp 489
QY 1699 CTTGTGTTGCGCTCCTCCCAAGGAGTATGACGATGAGCTGACCTTCTTCTCAAG 1758
Db 490 Leu-----ProLys-----GlyGluGluValThrIleLeuAlaGln 501
QY 1759 TCCAAAGGGGAACCAAAATTCACGCGTTAGGAACAGTGAAGTCCGCCGCCCTCGT 1818
Db 502 LysLysLys-----AspValTyrArgArgIleVal 511
QY 1819 AACACAAGCCTCGACACGCTTGAG---AGAGGCCACATGACACACACCATGGCAT 1875
Db 512 GluSerAspValGlyAspSerPheTyrIleArgThrHisPheGluTyrGluLysGluSer 531
QY 1876 CTTGGGACCTGGAATCTATCACCCAGGAATCTCAAACTCCCTTTGGCCCTGAACACGAG 1935
Db 532 ProTyrGly-----LeuSerPheAsnLysGly 540
QY 1936 CAGATAGGAACAGCTCGGCCACTTTTGAAGGCCAATGTGGAGGAAGGAGGAGCAGCC 1995
Db 541 GluValPheArgValValAspThrLeuTyrAsnGlyLysLeu---GlySerTrpLeuAla 559
QY 1996 AGCCGTTTGGGAGAGATCTCAAGGATCCAGACTCTCATCTTCCTTCTCT-----2043
Db 560 IleArgIleGlyLysAsnHisLysGluValGluArgGlyIleLeuProAsnLysAsnArg 579
QY 2044 -----CTGCGCCAGTGAATTTGTCTCTCCCA-----GCTTTGGGGGAG 2082
Db 580 AlaGluGlnLeuAlaSerValGlnTyrThrLeuProLysThrAlaGlyGlyAsp 597

RESULT 6
T30259
multiple PDZ domain protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30259
R:Simpson, E.H.; Suffolk, R.; Jackson, I.J.
Genomics 59, 102-104, 1999
A:Title: Identification, sequence, and mapping of mouse multiple PDZ domain protein 9
A:Residues: 1-2055 <SIM>
A:Molecule type: mRNA
A:Reference number: 220797; MUID:99326529; PMID:10395806
A:Accession: T30259
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Cross-references: EMBL:AJ131869; NID:g4150877; PIDN:CAA10523.1; PID:g4150878
A:Experimental source: strain C57/BL6 X CBA F1; whole brain
C:Genetics:
A:Gene: mpdz
A:Map position: 4

Alignment Scores:
Pred. No.: 2,01e-08 Length: 2055
Score: 259.00 Matches: 117
Percent Similarity: 35.93% Conservative: 72
Best Local Similarity: 22.24% Mismatches: 161
Query Match: 6.29% Indels: 176
DB: 2 Gaps: 19

US-09-502-945-4 (1-2236) x T30259 (1-2055)
```

```
QY 157 GCAGAGAGGACTATCTATGATGTGCGAATGTACCAACAGACCATGGACGTGGCC 216
Db 1673 AlaThrHisAspGluAlaIleAsnValLeuArgGlnThrProGlnArgVal-----1689
QY 217 GTGCTGCTGGGAGAGCTGAAGCTGGTCAATGAACCCAGCCGCTGCTGCTGTTGAT 276
Db 1690 -----ArgLeuThrLeuTyr---1694
QY 277 GCCATTGCGCGCTGATCCCACTGAAGCACCAGCTGGGAATATGATCAGCTACCCCGG 336
Db 1695 -----ArgAspGluAlaProTyrLysGluAspValCysAspThrPheThr-----1710
```


Db 373 AlaValAsnAsnValCysLeuGluGluValThrHisGluGluAlaValThrAlaLeuLys 392
 QY 562 ---ACCAAGAAACTGCTGCCATCAAGTG-----AGACACATC 597
 Db 393 AsnThrSerAspPheValTyrLeuLysValAlaLysProThrSerMetTyrMetAsnAsp 412
 QY 598 GGCCTGATCCCC-----GTGAAAAGCTCTCTGTATGAGCCCTCCTCATTGGCAGTATGTG 651
 Db 413 GlyTyrAlaProAspIleThrAsnSerSerGlnPro-----Val 427
 QY 652 GATCAGTTTGCTCGGAATCTGGGCGCTGCGAGGAGCCTGGGCTCCCTC----- 702
 Db 428 AspAsnHisValSerProSerPheLeuGlyGlnThrProAlaSerProAlaArgTyr 447
 QY 703 -----GGAATCGGGAA---AACAAAGGAGAAGGCTCTTC 735
 Db 448 SerProValSerLysAlaValLeuGlyAspAspGluIleThrArgGluProArgLysVal 467
 QY 736 ATCAGCCTGGTAGCTCCGAGGCTTGGCTGCAGATTTCCAGCGGCCCATCCACAGAG 795
 Db 468 ValLeuHisArgGlySerThrGlyLeuGlyPheAsnIleValGlyGlu---AspGly 486
 QY 796 CCTGGCATCTTATCAGCCATGTG-----AACTGGCTCCCTCTCTCTCCTGAGGTG 846
 Db 487 GluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu--- 505
 QY 847 GGATTGGAGATAGGACCAAGTGTTCGAAGTCAATGGCTGCTCTCTTAACCTGGAT 906
 Db 506 ---LeuArgLysGlyAspArgIleIleSerValAsnSerValAspLeuArgAlaAlaSer 524
 QY 907 CACAAGAGGCTGTAATGTCTGTAATAATAGC---CGCAGCCTGACCATC----- 954
 Db 525 HisGluGlnAlaAlaAlaLeuLysAsnAlaGlyGlnAlaValThrIleValAlaGln 544
 QY 954 ----- 954
 Db 545 TyrArgProGluGluTyrSerArgPheGluAlaLysIleHisAspLeuArgGluGlnMet 564
 QY 955 -----TCCATTGTAGCTGCAGCTGGC----- 975
 Db 565 MetAsnSerSerIleSerSerGlySerGlySerLeuArgThrSerGlnLysArgSerLeu 584
 QY 976 -----CGGAGCTGTTTCATGACAGACCGGCGCGCTGGCA----- 1011
 Db 585 TyrValArgAlaLeuPheAspTyrAspLysThrLysAspSerGlyLeuProSerGlnGly 604
 QY 1011 ----- 1011
 Db 605 LeuAsnPheLysPheGlyAspIleLeuHisValIleAsnAlaSerAspGluTrpTrp 624
 QY 1012 GAGCGCGCGCAG----- 1023
 Db 625 GlnAlaArgGlnValThrProAspGlyGluSerAspGluValGlyValIleProSerLys 644
 QY 1024 CTGAGCTGCAGCGGAGGAGCTTCTATGCGAAGCGGCTGGATGGAGTCCCAACAG 1083
 Db 645 ArgArgValGluLysLysGluArgAlaArgLeuLysThrValLysPheAsnSer---Lys 663
 QY 1084 ATCCTCCAGGAGCAGCAGAGATGGAGCGCAAGGAGAAAGAAATGCCCCAG----- 1137
 Db 664 ThrArgAspLysGlyGlnSerPheAsnAspLysArgLysLysAsnLeuPheSerArgLys 683
 QY 1138 -----AAGGCAGCAGGAGAAATGAGATACCGGAAAGGAGATGGAAACAGATT 1185
 Db 684 PheProPheTyrLysAsnLysAspGlnSerGlnGluThrSerAspAlaAspGlnHis 703
 QY 1186 GTA-----GAGGAGGAGAGAACTTTAAGAACGAATGGGAGAGACTGG 1230
 Db 704 ValThrSerAsnAlaSerAspSerGluSerSerTyrArg----- 716
 QY 1231 GGCTCAAGCAACAGTACTCTTGGCTTAAACCATCACTCTCTGAG-----GTA 1278
 Db 717 GlyGlnGluGlyTyrValLeuSerTyrGluProValAsnGlnGlnGluValAsnTyrThr 736

QY 1279 CACCAGATFACCCCTTCGCAAGCCAAAGTATGATCAGGAGTGGAACCTGAGCTCGAGCCC 1338
 Db 737 ArgProValIleIleLeuGlyProMetLysAspArgIleAsnAspAspLeuIleSerGlu 756
 QY 1339 GCAGATCACCTGGATGGAGGACCGGAGGAGCAGGAGGAGGATTTCCGGAATATGAG 1398
 Db 757 PheProAspLysPheGlySerCysValProHisThrThrArgProLysArgAspTyrGlu 776
 QY 1399 GAAGGCTTTGACCCCTACTCTATGTTTACCCAGCAGAGATCATCGGGAAGGATGTCGG 1458
 Db 777 ValAspGlyArgAspTyrHisPheValThrSerArgGluGlnMetGluLysAspIleGln 796
 QY 1459 CTCCTACGATCAAGAGGAGGATTCCTTACACCTGCCCTCGAAGGCGGTGGACTCC 1518
 Db 797 GluHisLysPheIleGluAlaGlyGlnTyrAsnAsnHisLeuTyrGly----- 812
 QY 1519 CCCATTGGGAAGGTGCTGTTCTGCTGTATGAGCGGGAGCTGTGAGCGGCATGGT 1578
 Db 813 -----ThrSerValGlnSerValArgGluValAlaGlyLysGlyLysHisCys 828
 QY 1579 GGCATTGTGAAGGAGCAGATCATGCAATCAACGCAAGATTTGTACAGACTACACC 1638
 Db 829 -----IleLeuAspValSerGlyAsnAlaIleLysArgLeuGln 841
 QY 1639 CTGGCTGAG 1647
 Db 842 IleAlaGln 844
 RESULT 8
 I38757
 homolog of Drosophila discs large protein, isoform 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
 C:Accession: I38757
 R:Luc, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.
 Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
 A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila
 A:Reference number: I38756; MUID:95024052; PMID:7937897
 A:Accession: I38757
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-904 <RES>
 A:Cross-references: EMBL:U13897; NID:g558437; PIDN:AAA50599.1; PID:g558438
 C:Genetics:
 A:Gene: GDB:DLG1
 A:Cross-references: GDB:393278; OMIM:601014
 A:Map position: 3q29-3q29
 C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
 C:Keywords: alternative splicing; duplication
 F:229-307/Domain: GLGF domain homology <GLG1>
 F:324-402/Domain: GLGF domain homology <GLG2>
 F:588-646/Domain: SH3 homology <SH3>
 F:715-892/Domain: guanylate kinase homology <GKI>

Alignment Scores:
 Pred. No.: 1,26e-07 Length: 904
 Score: 244.50 Matches: 121
 Percent Similarity: 38.88% Conservative: 87
 Best Local Similarity: 22.62% Mismatches: 202
 Query Match: 5.93% Indels: 125
 DB: 2 Gaps: 22

US-09-502-945-4 (1-2236) x I38757 (1-904)

QY 340 TCCAGGAGCTGAAGGAGTGGCTGTGACCGCTCTGCACCCGAGGCTCGGCCTGAGT 399
 Db 314 SerGluLysIleMetGluIleLysLeuLys---GlyProLysGlyLeuGlyPheSer 332
 QY 400 GTGCGCTGGTGGC-----CTGGAGTTTGGCTGTGGCTCTTCATCTCCAC 444
 Db 333 IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 352

Db	703	TyrGluProValAsnGlnGluValAsnTyrThrArgProValIleIleLeuGlyPro	722
	
Qy	1303	AAGTATGATCAGGAGTGGAACTGAGCTGAGCCGCCGAGATGACCTGGATGGAGGACGC	1362
		...	
Db	723	MetLysAspArgIleAsnAspLeuIleSerGluPheProAspLysPheGlySerCys	742
		
Qy	1363	GAGGAGCAGGAGAGCAGGATTTCCGGAAATATAGGAAGCTTTGACCCCTACTCTATG	1422
		
Db	743	ValProHisThrArgProLysArgAspTyrGluValAspGlyArgAspTyrHisPhe	762
		
Qy	1423	TTTCAACCCAGCAGCATCATCTGGGAGGATGTCGGCTCTACCATCAAGAAGGAGGGA	1482
		
Db	763	ValThrSerArgGluGlnMetGluLysAspIleGlnGluHisLysPheIleGluAlaGly	782
		
Qy	1483	TCCTTAGACCTGGCCCTGGNAGCGGTGTGGACTCCCCCATTTGGGAGGTGGCTTTCT	1542
		
Db	783	GlnTyrAsnAsnHisLeuTyrGly	794
		
Qy	1543	GCTGTGTATCAGCGGGGAGCTGTCGAGCGCATGCTGGCATTGTGAAAGGGGACGATC	1602
		
Db	795	SerValargGluValAlaGlyLysGlyLysHisCys	807
		
Qy	1603	ATGCCAATCAACGCAAGATTTGTGACAGACTACACCTGGGTGTAG	1647
		
Db	808	LeuAspValSerGlyAsnAlaIleLysArgLeuGlnIleAlaGln	822
		
RESULT 9			
T23160			
Hypothetical protein K01A6.1 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000			
C:Accession: T23160			
R:Cottage, A.			
submitted to the EMBL Data Library, January 1996			
A:Reference number: Z19701			
A:Accession: T23160			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-1012 <WIL>			
A:Cross-references: EMBL:Z68750; PIDN:CAA92965.1; GSPDB:GN00022; CESP:K01A6			
A:Experimental source: clone K01A6			
C:Genetics:			
A:Gene: CESP:K01A6.1			
A:Map position: 4			
A:Introns: 107/1; 151/2; 183/2; 270/2; 316/3; 410/3; 628/3; 922/3; 971/3			
C:Superfamily: WW repeat homology			
F:131-168/Domain: WW repeat homology <WWR>			
Alignment Scores:			
Pred. No.:	2,04e-07	Length:	1012
Score:	241.50	Matches:	121
Percent Similarity:	34.84%	Conservative:	72
Best Local Similarity:	21.84%	Mismatches:	194
Query Match:	5.86%	Indels:	167
DB:	2	Gaps:	19
US-09-502-945-4 (1-2236) x T23160 (1-1012)			
Qy	29	CCAGTCTCTGGCAGCGGGACCCGAGGAAGCGGTCTG	76
Db	574	ProLysThrArgSerArgThrProSerAlaAlaPheArgTyrGlyGluProGlnThrAsn	593
		
Qy	77	CAGCTGGACCTGCCCCAGCCA	103
		
Db	594	MetMetAspSerAlaAlaProLeuProValArgSerLysThrProAlaGluArgGlnThr	613
		
Qy	104	GAAAGTGGCCGAGAAATTCGGCATAAGTGGATTTTCTGATTGAAAAATGATGCAGAGA	163
		
Db	614	SerArgThrGluGluAspGlnAsnValArg	628
		
Qy	164	AGGACTATCTCTATGATGTGTCGGAATGTACCAACGACCATGGCGCTGGCTGCTCG	223
		

Db 629 Gln-----ProAlaValThrSerGluTrp-GluGlyMetSerSe 641
Qy 224 TGGGAGACCTGAAGTGTCTCATCAATCAACCCAGCGCTGCTCTGTTTATGTCGCAATC 283
Db 641 AlaIlePro---AlaSerArgMet-ArgProSerSerThrThrLeuGlyPheAlaThrP 660
Qy 284 GCCCCTGATCCACTGAAG-----CACAGGTGGATATGATCAGTGCACCCCGGC 337
Db 660 toAsnTyIleProLeuSerGlnTyAsnGlnLysProSerAspLeuIleThrValSerL 680
Qy 338 GCTCCAGGAAGCTGAAGGAGTGTCTGGACCGTCTCCACCCGAGCGCTCCGCTGA 397
Db 680 euIleArgLys-----ProValGlyPheGlyPheA 690
Qy 398 GTGTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 457
Db 690 rgLeuLeuGlyGlyValGluSerLysThrProLeuSerValGlyGlnIleValIleGlyG 710
Qy 458 GTCAGGCAGACGCTCGGG---CTCCAGGTAGGGAGGAGATCCTCCGGATCAATGAT 514
Db 710 lYAlaAlaGluAspGlyArgLeuGlnGluGlyAspGluIleValGluIleAspGlyH 730
Qy 515 APTCATCTCCCTGTCACCATGAGGAGTCAATCAACTCATTCGAACCAAGAAACTG 574
Db 730 isAsnValGluClyAlaSerHisSerGluAlaValValLeuLeuGlu-----A 746
Qy 575 TGTCCATCAAGTGAACATCGCCCTGATCCCGTGAAGAGTCTCTCTGATGAGCC 634
Db 746 laAlaAlaGlnAsnLysHisValLysLeuIleValArgArgProSerArgThrAspProA 766
Qy 635 TCACITGGCAGTATGTGGATCAGTTGTGCGAATCTGGGGCGTGGCGGACGCTGG 694
Db 766 laArg-----ArgGlySerLeuA 772
Qy 695 GTCCTCCCTGGAATCGGAAACAAAGGAGAGAGTCTTCACTGAGCTGGTGGTGGTGG 754
Db 772 snSerAlaGlyProSerGlySerTyAsp-----ValLeuLeuHisArgAsnGluAsnA 790
Qy 755 GAGGCTTGGTGGCAGATTTCCAGCGGCCCATCCAGAGGCTGGCATCTTTATCAGCC 814
Db 790 spGlyPheGlyPheValLeuMetSerSerGln---HisLysAsnGlySerThrValGlyG 809
Qy 815 ANGTGAAACCTGCCCTCTGCTGAGTGGGA---TTGAGATAGGGGACGACGATG 871
Db 809 InIleGlnProGlySerProAlaAlaArgCysGlyArgLeuSerValGlyAspArgValI 829
Qy 872 TCGAAGTCAATGGCTGCTACTTCTTAACCTGGATCAAGAGCGCTGTAATGTCTGA 931
Db 829 leAlaValAsnGlyIleAspIleLeuSerLeuSerHisProAspThrIleSerLeuIleL 849
Qy 932 AAATAGCCGC-----AGCCTGACATCTCCATTTGATGCTGCGCGCGGAGC 982
Db 849 ysAspSerGlyLeuSerValArgLeuThrIleAlaProAsnThrAlaGlyProValI 869
Qy 983 TGTTCATGACACCGGGAGCGGTGGCAGAGCGCGGCGGCTGAGCTGCAGCGCGCAGG 1042
Db 869 euProMetVal-----S 873
Qy 1043 AGCTTCTCATGACAGAGCGGTGGCGATGGAGTCCAAAGATCCTCCAGGAGCAGCAGG 1102
Db 873 erAlaThrLeuGlyArgAsnPheThrMetAsnGlyHis----- 885
Qy 1103 AGATGAGCGGCAAGGAGAAAGAAATTTGCCAGAGGACGACAGAGAAATGAGAT 1162
Db 885 ----- 885
Qy 1163 ACCGGAAGGAGATGGACAGATTTAGAGGAGAGAGAGATTTAAGAACCAATGGGAG 1222
Db 886 -----Tyrlus 888
Qy 1223 AAGACTGGGCTCAAGGAACAGCTACTCTTGCCTAAACCATCACTGCTGAGGTACACC 1282
Db 888 erAsnTyrlGly-----LeuProProProProSerValTyrlGluLysHisP 904

Qy 1283 CAGTACCCCTCCGCAAGCCAAAGTATGATCAGGAGTGAACCTGAGCTCGAGCCCGCAG 1342
Db 904 roProProTyrlLeuAlaPheAspGlyLeuSerIleAsnAspArgMetSerMetAsnG 924
Qy 1343 ATGACCTGGATGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1402
Db 924 lyAsnLeuIleAspValThrLeuGluArgGlyThrLysGlyPhe-----G 939
Qy 1403 GCTTTGACCCCTACTCTATGTTTCCACCCAGAGCAGATCATGGGGAAGGATGTCGCGCTCC 1462
Db 939 lyPhe----- 940
Qy 1463 TACGCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAGCGGCTGGACTCCCGCCA 1522
Db 941 -----SerIleArgGlyGlnGluPheGlyS 950
Qy 1523 TTGGGAAGTGGTGGTTCCTGCTGTGTATGAGCGGGAGCTGTGAGCGGCTGGTGGCA 1582
Db 950 erMetProLeuPheValLeuArgIleAlaAspGlyProAlaLysAlaAspGlyArgL 970
Qy 1583 TTGTGAAGGGGAGGAGATCATGCGCAATCAACGGCAAG 1620
Db 970 euGlnValGlyAspGlnLeuThrThrIleAsnGlyGln 982
RESULT 10
I56552
synapse-associated protein 97 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jan-2000
C:Accession: I56552
J:Mueller, B.M.; Kistner, U.; Veh, R.W.; Cases-Langhoff, C.; Becker, B.; Gundelfinger, R. Neurosci. 15, 2354-2366, 1995
A:Title: Molecular characterization and spatial distribution of SAP97, a novel presyn A:Reference number: I56552; MUID:95198112; PMID:7891172
A:Accession: I56552
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-911 <RES>
A:Cross-references: EMBL:U14950; NID:9642455; PIDN:AAA79976.1; PID:9642456
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h F:229-307/Domain: GLGF domain homology <GLGI>
F:323-401/Domain: GLGF domain homology <GLG2>
F:587-645/Domain: SH3 homology <SH3>
F:722-899/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 2,27e-07 Length: 911
Score: 240.50 Matches: 119
Percent Similarity: 38.31% Conservative: 89
Best Local Similarity: 21.92% Mismatches: 200
Query Match: 5.84% Indels: 135
DB: 2 Gaps: 22
US-09-502-945-4 (1-2236) x I56552 (1-911)
Qy 343 AGGAAGTGAAGGAGGTCGCTGTCGACCGCTGCACCCGAGCGCTCGGCTGAGTGTG 402
Db 314 ArgLysAsnHisGluIleLysLeuIleLys---GlyProLysGlyLeuGlyPheSerile 332
Qy 403 CGTGGTGGC-----CTGGAGTTTGGCTGTCCTTCATCTCCACCTC 447
Db 333 AlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrlValThrLysile 352
Qy 448 ATCAAGCGCGTTCAGGACGACAGCGCTCGGG---CTCCAGGTAGGAGCGAGATCGTCCGG 504
Db 353 lIleGluGlyGlyAlaAlaHisLysAspGlyLysLeuGlnIleGlyAspLysLeuLeuAla 372
Qy 505 ATCAATGATATTCATCTCTCTCTGATACCATGAGGAGGTCAATCAACCTCATTCGA--- 561
Db 373 ValAsnSerValCysLeuGluGluValThrHisGluAlaValThrAlaLeuLysAsn 392
Qy 562 ACCAAGAAACTGTGTCATCAAGTGAGA-----CACATCGGC 600

```

Db 393 ThrSerAspPheValTyrLeuLysAlaAlaLysProThrSerMetTyrIleAsnAspGly 412
Qy 601 CTGATCCCGTGAAGCTCTCTGAT---GAGCCCTCAGTGGCAGATGTGTGATCAG 657
Db 413 TyrAlaPro-----ProAspIleThrAsnSerSerGlnSerValAspAsn 428
Qy 658 TTTGTGCGGAATCTGGGGCTGCGAGGCGCTGGGCTGCCCT-----702
Db 429 HisValSerProSerSerTyrLeuGlnThrProAlaSerProAlaArgTyrSerPro 448
Qy 703 -----GGAATCGGAA-----AACAAGGAGGAAGAGTCTTCATCAGC 741
Db 449 IleSerLysAlaValLeuGlyAspAspGluIleThrArgGluProArgLysValValLeu 468
Qy 742 CTGATAGCTCCGAGGCTGGCTGCAGCATTTCCAGCGGCCCATCCAGAGCCTGGC 801
Db 469 HisArgGlySerThrGlyLeuGlyPheAsnIleValGlyGlu---AspGlyGluGly 487
Qy 802 ATCTTTATCAGCATGTG-----AAACTGGCTCCCTGCTGCTGAGGTGGATG 852
Db 488 IlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu-----Leu 505
Qy 853 GAGATAGGGCCAGCATGTGCAAGTCAATGCGCTGCTCTTAACCTGGATCACAAG 912
Db 506 ArgLysGlyAspArgIleIleSerValAsnSerValAspLeuArgAlaAlaSerHisGlu 525
Qy 913 GAGGTGTAATGTCTGAAATAAGCCGACCTGACCATCTCCATTTAGCTGACGCT 972
Db 526 GlnAlaAlaAlaLeuLysAsnAla---GlyGlnAlaValThrIleValAlaGlnTyr 544
Qy 973 GCCCGGAGCTTTCATGACAGACCGGAGCGGCTGGCGAGAGCGCGGAG-----1023
Db 545 ArgProGluGluTyrSerArgPheGluAlaLysIleHisAspLeuArgGluThrMetMet 564
Qy 1024 -----CCTGAGCTGCAGCGCAGGAGCTTCTC 1050
Db 565 AsnSerSerValSerSerGlySerGlySerLeuArgThrSerGlnLysArgSerLeuTyr 584
Qy 1051 ATGCAGAGCGGCTGGCGATGAGTCCACAAG-----1083
Db 585 ValArgAlaLeuPheAspTyrAspLysThrLysAspSerGlyLeuProSerGlnGlyLeu 604
Qy 1083 -----1083
Db 605 AsnPheLysPheGlyAspIleLeuHisValIleAsnAlaSerAspAspGluTrpTrpGln 624
Qy 1084 -----ATCCTCCAGGAGCAGCAG 1101
Db 625 AlaArgGlnValThrProAspGlyGluSerAspGluValGlyValIleProSerLysArg 644
Qy 1102 GAGATGAGCGGCAAGGAGAAAGAAATGCCAGAGGAGCAGAGAGAAATGAGAGA 1161
Db 645 ArgValGluLysLysGluAlaArgLeu-----LysThrValLysPheAsnSerLys 662
Qy 1162 TACCGAGGAGATGACAGATTTAGAGAG-----1194
Db 663 ThrArgGlyAspLysGlyGluIleProAspAspMetGlySerLysGlyLeuLysHisVal 682
Qy 1195 -----GAAGAGAAGTTTAAGAGCAATGGAGAGACTGGGGC 1233
Db 683 ThrSerAsnAlaSerAspSerGluSerSerTyrHisGluTyrGlyCysSerLysGlyGly 702
Qy 1234 TCAAGGAGAACGCTACTCTTGCTTAAACCATCATCTGCTGAG-----GTACAC 1281
Db 703 GlnGluGluTyrValLeuSerTyrGluProValAsnGlnGlnGluValAsnTyrThrArg 722
Qy 1282 CAGTACCCCTTCGCAACCCAAAGTATGATCAGGAGGTGGAACTGAGCTCAGCCCCCA 1341
Db 723 ProValIleIleLeuGlyProMetLysAspArgValAsnAspAspLeuIleSerGluPhe 742
Qy 1342 GATGACCTGGATGAGGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1401

```

```

Db 743 ProAspLysPheGlySerCysValProHisThrThrArgProLysArgAspTyrGluVal 762
Qy 1402 GCCTTTGACCCCTACTCTATGTTCCACCCAGAGCAGATCATGGGGAAGGATGTCCGGCTC 1461
Db 763 AspGlyArgAspTyrHisPheValThrSerArgGluGlnMetGluLysAspIleGlnGlu 782
Qy 1462 CTACGCATCAAGAGGAGGATCCTTAGACCTGCGCCCTGGAGGC---GGTGTGACTCC 1518
Db 783 HisLysPheIleGluAlaGlyGlnTyrAsnAsnHisLeuTyrGlyThrSerValGlnSer 802
Qy 1519 CCCATTGGGAAGGTGCTGCTGCTGCTGATGAGCGGGGAGCTGTCGAGCGCATGGT 1578
Db 803 -----ValArgAlaValAlaGluLysGly-----LysHisCys 813
Qy 1579 GGCATTGTGAAGGGGAGCAGATCATGGCAATCAACGCGAGATGTGACAGACTACACC 1638
Db 814 -----IleLeuAspValSerGlyAsnAlaIleLysArgLeuGln 826
Qy 1639 CTGGCTGAG 1647
Db 827 IleAlaGln 829

RESULT 11
G01974
channel associated protein of synapse - human
C/Species: Homo sapiens (man)
C/Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 21-Jan-2000
C/Accession: G01974
R/Kim, E.; Cho, K.
submitted to the EMBL Data Library, July 1995
A/Reference number: G08966
A/Accession: G01974
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: mRNA
A/Residues: 1-870 <KIM>
A/Cross-references: EMBL:U32376; NID:g1463025; PID:g1036790
C/Genetics:
C/Gene: chapsyn-110
C/Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
F:198-276/Domain: GLGF domain homology <GLG2>
F:543-601/Domain: SH3 homology <SH3>
F:681-858/Domain: guanylate kinase homology <GKI>

Alignment Scores:
Pred. No.: 1.04e-06 Length: 870
Score: 230.00 Matches: 143
Percent Similarity: 35.74% Conservative: 114
Best Local Similarity: 19.89% Mismatches: 222
Query Match: 5.58% Indels: 240
DB: 2 Gaps: 32

US-09-502-945-4 (1-2236) x G01974 (1-870)
Qy 208 GACGTGGCCGTGCTGCTGGGAGACCTGAAGCTGCTCATCAATGAA-----CCCAGCCGT 261
Db 141 AspGlyArgLeuArgValAsnAspCysIleLeuArgValAsnGluValAspValSerGlu 160
Qy 262 CTGCTCTGTTGATGATTCATTCCTGCGCCGCTGATCCCTGAGCAGCACCAGGTGGAATATGAT 321
Db 161 ValSerHisSerLysAlaValGlu-----AlaLeuLysGluAlaGlySerIleAla 177
Qy 322 CAGCTGACCCCGCGCTCCAGGAAGCTG-----AAGGAGGTGCGTCTGTGAC 369
Db 178 ArgLeuTyrValArgArgArgProIleLeuGluThrValValGluIleLysLeuPhe 197
Qy 370 CGTCTGACCCCGAGCCCTCGGCTGAGTGTGCGTGGTGGC-----CTG 414
Db 198 Lys---GlyProLysGlyLeuGlyPheSerIleAlaGlyValGlyAsnGlnHisIle 216
Qy 415 GAGTTTGGCTGGGCTTTCATCTCCACCTCATCAAGCGGTGACGAGCAGCAGCGCT 474
Db 217 ProGlyAspAsnSerIleTyrValThrLysIleLeuAspGlyAlaAlaGlnLysAsp 236

```

QY	475	GGG---CTCCAGGTTAGGGACAGATCGTCCGGATCAATCGATATTCCATCTCCTCCTGT	531
DB	237	GLYArgLeuGlnValGlyAspArgLeuLeuMetValAsnAsnTyrSerLeuGluGluVal	256
QY	532	ACCCATGAGGAGGTCACTCACTCATTCGA---ACCAAGAAACTGTGCCATCAAGNG	588
DB	257	ThrHisGluGluAlaValAlaIleuLeuLysAsnThrSerGluValValTyrLeuLysVal	276
QY	588	-----	588
DB	277	GlyAsnProThrThrIleTyrMetThrAspProTyrGlyProProAspIleThrHisSer	296
QY	589	-----AGACACATC-----GGCTGATCCCGGTGAAA	615
DB	297	TyrSerProMetGluAsnHisLeuLeuSerGlyAsnAsnGlyThrLeuGluTyrLys	316
QY	616	AGCTCT- ::: ::	621
DB	317	ThrSerLeuProIleSerProGlyArgTyrSerProIleProLysHisMetLeuVal	336
QY	622	-----CCTGATGAGCCCTCACTTGGCAGTAGTATGGTACAGTTTGTG	663
DB	337	AspAspTyrThrArgProProGluProVal---TyrSerThrValAsnLysLeuCys	355
QY	564	TCGGATCTGGGGCGTGGCA-----	584
DB	356	AspLysProAlaSerProArgHisTyrSerProValGluCysAspLysSerPheLeuLeu	375
QY	685	-----GCG	687
DB	376	SerAlaProTyrSerHisTyrHisLeuGlyLeuLeuProAspSerGluMetThrSerHis	395
QY	688	AGCCTGGCTCCCTCGAAATCGG-----	711
DB	396	SerGlnHisSerThrAlaThrArgGlnProSerMetThrLeuGlnArgAlaValSerLeu	415
QY	712	GAACAACAGGAGAGAAGGTCTTCATCAGCCTGGTAGCTCCCGAGCCCTTGCTGCAGC	771
DB	416	GluGlyGluProArgLysValValLeuHis---LysGlySerThrGlyLeuGlyPheAsn	434
QY	772	ATTTCACAGCGGCCCATCCAGAAGCTGGCATCTTATCACCATGTG-----AAA	822
DB	435	IleValGlyGlyGlu---AspGlyGluGlyIlePheValSerPheIleLeuAlaGlyGly	453
QY	823	CCTGGCTCCCTGTCTGCTGAGTGGGATGAGATPAGGGACACAGATGTGCAAGTCAAT	882
DB	454	ProAlaAspLeuSerGlyGlu-----LeuGlnArgGlyAspGlnIleLeuSerValAsn	471
QY	883	GGCGTGCATCTTCTTAACCTGGATCACAAGGAGCGTGTAAATGTCTGAAATAATGCCGC	942
DB	472	GlyIleAspLeuArgGlyAlaSerHisGluGlnAlaAlaAlaLeuLysGlyAla---	490
QY	943	AGCCTGACCATCTCACTTACGTACGTGCGCGGAGCTGTTTCATGACAGACCGGGAG	1002
DB	491	GlyGlnThrValThrIleIleAlaGlnTyrGlnProGluAspTyrAlaArgPheGluAla	510
QY	1003	CGGTGGCAGACGCGCGGCAG-----	1023
DB	511	LysIleHisAspLeuArgGluGlnMetMetAsnHisSerMetSerSerGlySerGlySer	530
QY	1024	---CGTGAGCTGACGGCAGGAGCTTCTCATCGAC-----	1056
DB	531	LeuArgThrAsnGlnLysArgSerLeuTyrValArgAlaMetPheAspTyrAspLysSer	550
QY	1056	-----	1056
DB	551	LysAspSerGlyLeuProSerGlnGlyLeuSerPheLysTyrGlyAspIleLeuHisVal	570
QY	1057	-----AAGCGCTGGCGATGGATGCTCCACAAG	1083
DB	571	IleAsnAlaSerAspAspGluTrpTrpGlnAlaAlaArgValMetLeuGluGlyAspSer	590
QY	1084	-----ATCCTCCAGGACGACGAGGAGATGGAGCGCGCAAGAGAAAGAAAGAA	1128

[illegible]

Percent Similarity:	34.30%	Conservative:	86
Best Local Similarity:	20.39%	Mismatches:	209
Query Match:	5.52%	Indels:	197
DB:	2	Gaps:	22
US-09-502-945-4 (1-2236) x T15617 (1-1131)			
Qy	151	AATGATGCAGAGAAGGACTATCTCTATGATGTGTCGGAATGTACCAACAGACCATGGAC	210
Db	243	AsnSerSerThrHisTyrLeuHisGluArgGlnArg-----GlnThrSerHis	259
Qy	211	GTGGCCGTGCTCGGGAGACCTGAAGCTGGTCATCAAGAACCCAGCCGTCTGCCTCTG	270
Db	260	AspGlyThrTrpArgGluThrThrThrArgThrValAspThrProSerGlyLeuGlu---	278
Qy	271	TTTGATGCATTCGGCCGCTGATCCCACTCAAGCACAGGTGGAAATATGATCAGCTGACC	330
Db	279	-----ArgArgValValGluHisThrGlyValle	288
Qy	331	CCCGGGCGCTCCAGAGAG-----CTGAAGAGGTGGCTGTGGACCGTCTGCACCCCGAA	384
Db	289	AspAspHisGlyArgLysTrpGluLeuGluAsnIleValLeuGluLysGlyHis---Thr	307
Qy	385	GGCTCGGCCCTGAGTCGCTGGTGGCCG-----GAGTTTGGCTGTGGG	429
Db	308	GlyLeuGlyPheSerIleThrGlyGlyMetAspGlnProThrGluAspGlyAspThrSer	327
Qy	430	CTCTTCATCTCCCACTCATCAAGCGGTGACGACAGAGCGTCCGG---CTCCAGGTA	486
Db	328	IleYrValThrAsnIleleGluGlyGlyAlaAlaLeuAlaAspGlyArgMetArgLys	347
Qy	487	GGGACGAGATCGTCCGGATCAATGATATTCATCTCTCTGTACCATGAGGAGTGC	546
Db	348	AsnAspIleIleThrAlaValAsnAsnThrAsnCysGluAsnValLysIleGluValAla	367
Qy	547	ATCAACCTCATFTCCAAACCAAGAAACTGTG-----TCCATCAAGTGAGA-----	591
Db	368	ValAsnAlaLeuLysSerSerGlyAsnValSerLeuSerLeuLysArgArgLysAsp	387
Qy	591	-----GluAlaPheLeuProIleGlyGlyAsnPheGlyGlySerThrSerTyrLeuArgSerGly	407
Db	592	-----CACATCGCCCTGATCCCC	609
Db	408	ValThrProSerValSerAlaGlyAsnLeuGlnHisAlaIleHisSerProSerAlaPro	427
Qy	610	GTGAAAGCTCTCCTGATGAGCCCTCCTGATGTCAGTATGTGGATGTGATCTGTCTGCGAA	669
Db	428	IleHisProProProProProValHisHisGlySerLeuSerGlnLeu-----	444
Qy	670	TCTGGGGCGTGGAGCGACCTGGGCTCCCTCGGAATTCGGAAACAGGAGGAAG	729
Db	445	-----SerValGlyGln-----TyrArgSerThrArgProAsnThr	456
Qy	730	GTCATTCACTACCGCTGTA---GGCTCCGAGGCTTGGCTGGCAGATATTCACGCGCCCC	786
Db	457	SerValleAspLeuValLysGlyAlaArgGlyLeuGlyPheSerIleAlaGlyGln	476
Qy	787	ATCCAGAAG-----CCTGGCATCTTTATCAGCCAT---GTGAAACCTGCC	828
Db	477	GlyAsnGluHisValLysGlyAspThrAspIleTyrValThrLysIleleGluGluGly	496
Qy	829	TCCCTGTCTGCTGAGTGGGATGGAGATAGGGGACCATGATGTGGAAGTCAATGGCTC	888
Db	497	AlaAlaGluLeuAspGlyArgLeuArgValGlyAspLysIleLeuGluValAspHis	516
Qy	889	GACTTCTCTAACCTGGATCAAGAGGCTGTAAATGTCTGAAAAATAGCGCAGCTG	948
Db	517	SerLeuIleAsnThrThrHisGluAsnAlaValAsnValLeuLysAsnThrGlyAsn---	535
Qy	949	ACCATCTCCATTGTAGCTGCAGCTGGCCGGAGCTGTTTCATGACAGACGGGAGCGGTG	1008


```

Db 761 AsnAlaIleGlyArgLeuGlnValAlaGlnLeuTyrProIleAlaIlePheIleLysPro 780
QY 1594 -----GACGAGATCATGCAATCAACGGCAGATTGTGACAGACTACACCTGGCT 1644
Db 781 LysSerLeuGluProLeuMetGluMetAsnLysArgLeuThrGluGlnAlaLysLys 800
QY 1645 GAGGCTGACGCTGCCCTGAGAGGCTGGATCAGGCGGGGAGCTGGATCGACCTTGG 1704
Db 801 ThrTyrAspArgAlaIle---LysLeuGluGlnGluPheGlyGlyTyrPheThrAlaIle 819
QY 1705 GTT 1707
Db 820 Val 820
RESULT 14
S71625
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 13 - mouse
N:Alternate names: epidermal growth factor-binding protein; serine proteinase
C:Species: Mus musculus (house mouse)
C:Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 21-Jun-2002
C:Accession: S71625; S67987; I81210; I81209; S40290
R:Chida, D.; Kume, T.; Makouyama, Y.; Tabata, S.; Nomura, N.; Thomas, M.L.; Watanabe, T.
PDBS Lett. 358, 233-239, 1995
A:Title: Characterization of a protein tyrosine phosphatase (RIP) expressed at a very ea
A:Reference number: S71625; MUID:95145716; PMID:7843407
A:Accession: S71625
A:Molecule type: mRNA
A:Residues: 1-2450 <CHI>
A:Cross-references: EMBL:D83966; NID:g1232103; PIDN:BAAL12158.1; PID:g1232104
A:Experimental source: strain DBA/2; cell line MEL 745A
R:Wolf, B.B.; Brown, M.D.
PDBS Lett. 376, 177-180, 1995
A:Title: Epidermal growth factor-binding protein activates soluble and receptor-bound si
A:Reference number: S67987; MUID:96105375; PMID:7498536
A:Accession: S67987
A:Molecule type: protein
A:Residues: 1098-1102 <WOL>
A:Experimental source: submaxillary glands
R:Sato, T.; Irie, S.; Kitada, S.; Reed, J.C.
Science 268, 411-415, 1995
A:Title: FAP-1: a protein tyrosine phosphatase that associates with Fas.
A:Reference number: I59595; MUID:95232528; PMID:7536343
A:Accession: I81210
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1329-1354, 'K', 1356-1447, 'R', 1449-1454 <RES>
A:Cross-references: GB:L34582; NID:g806297; PIDN:AAC42056.1; PID:g806298
A:Accession: I81209
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1338-1354, 'K', 1356-1447, 'R', 1449-1454 <RE2>
A:Cross-references: GB:L34581; NID:g806295; PIDN:AAC42055.1; PID:g806296
R:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.
submitted to the EMBL Data Library, June 1993
A:Description: Assessment of the expression levels of murine protein-tyrosine phosphatas
A:Reference number: S40280
A:Accession: S40290
A:Molecule type: mRNA
A:Residues: 2266-2372 <HEN>
A:Cross-references: EMBL:D23059; NID:g438155; PIDN:CAA80594.1; PID:g438156
C:Genetics:
A:Gene: Ptpn13
A:Map position: 5
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology;
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; tyros
F:566-860/Domain: protein 4.1 membrane-binding domain homology <B41>
F:1089-1165/Domain: GLGF domain homology <GLG1>
F:1361-1437/Domain: GLGF domain homology <GLG2>
F:1495-1574/Domain: GLGF domain homology <GLG3>
F:1769-1840/Domain: GLGF domain homology <GLG4>
F:1863-1937/Domain: GLGF domain homology <GLG5>
F:2203-2422/Domain: protein-tyrosine-phosphatase homology <PTP1>

```

```

F:2374/Active site: Cys (phosphocysteine intermediate) #status predicted
F:2380/Binding site: substrate phosphate (Arg) #status predicted

Alignment Scores:
Pred. No.: 5,19e-06 Length: 2450
Score: 221.50 Matches: 144
Percent Similarity: 30.88% Conservative: 104
Best Local Similarity: 17.93% Mismatches: 356
Query Match: 5.38% Indels: 2
Gaps: 2
DB:

```

US-09-502-945-4 (1-2236) x S71625 (1-2450)

```

QY 308 AGGTGGAATATGATCAGCTGACCCCGCGCTCAGGAAGCTAAGAGGTGGCTGTGG 367
Db 1070 ArgTrpSerIleValSer--SerProGluArgGluIleThrLeuValAsnLeuLysLysA 1089
QY 368 ACCGTCCTCACCCGAGAGGCTCGGCTGAGTGTGCGGTGGCTGGAGTTTGGC--- 423
Db 1089 spProLysHis-----GlyLeuGlyPheGlnIleIleGlyGlyGlyLysMetGlyArgL 1107
QY 424 ----TGTGGCTCTCTCATCTCCACCTCATCAAGGCGGTGTCAGGACAGACGCTCGGG- 477
Db 1107 euAspLeuGlyValPheIleSerAlaValThrProGlyGlyProAlaAspLeuAspGlyC 1127
QY 478 --CTCCAGGTAGGGACAGATCGTCCGGATCAATGGATATTCATCTCCTCTGTACCC 535
Db 1127 ysLeuLysProGlyAspArgLeuIleSerValAsnSerValSerLeuGluGlyValSerH 1147
QY 536 ATGAGGAGGTATCAACCTCATTCGA----- 561
Db 1147 IsHisAlaAlaValAspIleLeuGlnAsnAlaProGluAspValThrLeuValIleSerG 1167
QY 561 ----- 561
Db 1167 lnProLysGluLysProSerLysValProSerThrProValHisPheAlaAsnGlyMetL 1187
QY 562 -----ACCAAGAAACTGTGCCATCAA----- 585
Db 1187 ysSerTyrThrLysLysProAlaTyrMetGlnAspSerAlaMetAspProSerGluAspG 1207
QY 586 -----GTGAGACACATCGGCTGATCCCTGATCCCGTGAAGAGCTCTCTG 625
Db 1207 lnProTrpProArgGlyThrLeuArgHisIle-----ProG 1219
QY 626 ATGAGCCCTCACTTGGCAGTATGTGGATCATGTTGTGCGAATCTGGGGCGCTGGCA- 684
Db 1219 luSerProPhe-----GlyLeuSerGlyGlyLeuArgG 1230
QY 685 --GGCAGCTGGCTCCCTCGAATATCGG-----GAAACAAAGG 721
Db 1230 luGlySerLeuSerSerGlnAspSerArgThrGluSerAlaSerLeuSerGlnSerGlnV 1250
QY 722 AGAAGAAGTCTTCATCATCAGCCTGTAGCTCCCGAGG----- 759
Db 1250 alaAsnGlyPhePheAlaSerHisLeuGlyAspArgGlyTrpGlnGluProGlnHisSerS 1270
QY 759 ----- 759
Db 1270 erProSerProSerValThrThrLysValAsnGluLysThrPheSerAspSerAsnArgS 1290
QY 759 ----- 759
Db 1290 erLysAlaLysArgGlyIleSerAspLeuIleGluHisLeuAspCysAlaAspSerA 1310
QY 759 ----- 759
Db 1310 spLysAspAspSerThrTyrThrSerSerGlnAspHisGlnThrSertysGlnGluProS 1330
QY 759 ----- 759
Db 1330 erSerSerLeuSerThrSerAsnLysThrSerPheProThrSerSerAlaSerProProL 1350

```

```
QY 760 -----CTTGCTGCAGCA 772
Db 1350 ysProGlyAspThrPheGluValGluLeuAlaLysThrAspGlySerLeuGlyIleSerV 1370
QY 773 TTTCCAGCGC-----CCATCCAGAGCGTGGCATCTTTATCAGCCATGTGAAC 823
Db 1370 alThrGlyGlyValAsnThrSerValArgHisGlyIleYrValLysAlaIleIleP 1390
QY 824 CT---GGCTCCCTGCTGCTGAGTGGGATGGAGATAGGGACACGATTGTGGAAGTCA 880
Db 1390 roLysGlyAlaAlaGluSerAspGlyArgIleHisLysGlyAspArgValLeuAlaValA 1410
QY 881 ATGGCTCGACACTCTCTAACCCTGGATCACAAAGAGGCTGTAAATGTGTGAATAATAGCC 940
Db 1410 snGlyValSerLeuGluGlyAlaThrHisLysGlnAlaValGluThrLeuArgAsnThrG 1430
QY 941 GCAGCCTGACCATCTCCATGTTAGCTGCAGCTGGCGGAGAGCTGTCATCAGACAGCCGG 1000
Db 1430 lyGlnValVal----- 1433
QY 1001 AGCGGCTGGCAGAGCGCGCGCAGCGTGAGCTGCAGCGGCGAGGAGCTCTCATGCAGAAC 1060
Db 1434 -----HisLeuLeuLeuGluLys- 1439
QY 1061 GGCTGCGATGGAGTCCAAACAAGATCCTCCAGAGCAGCAGGAGATGGAGCGCAAGGA 1120
Db 1440 -----GlyGlnValProThrSerArgGluGlnA 1449
QY 1181 AGATTGTAGAGGAGGAAGACAAGTTTAAGAAAGCAATGGGAAGAAAGACTGGGGCTCAAAG 1240
Db 1469 ysValAlaLysGlnThrProHisValLys-----AspTyr----- 1480
QY 1241 AACAGCTACTCTTGGCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGC 1300
Db 1481 -----SerPheValThrGluAspAsnThrPheGluValLysLeuP 1494
QY 1301 CAAAGTATGATCAGGAGTGGAACCTGAGCTGCAGCGCCGAGATGACCTGGATGGAGGCA 1360
Db 1494 heLysAsnSerSerGlyLeuGlyPheSerPheSerArgGluAspAsnLeu----- 1510
QY 1361 CGGAGGACGAGGAGACGAGGATTTCCGAAATATGAGGAAGGCTTTGACCCCTACTCTA 1420
Db 1510 ----- 1510
QY 1421 TGTTCACCCAGACAGATCATGGGAGGAGTCTCGGCTCTCCGCTCCTACGATCAAGAAGAGG 1480
Db 1511 -----lleProGluGlnIleAsnGlySer-----lleValArgValLysLys--- 1524
QY 1481 GATCCTTAGACTGGCCCTCGAAGCGGTGTGGACTCCCTCCCATTTGGGAAGGTGTCGTTT 1540
Db 1524 ----- 1524
QY 1541 CTGCTGTGTATAGCGGGAGCTGCTGAGCGGCGATGTGGCATTTGGATGAAAGGGAGCAGA 1600
Db 1525 -----LeuPheProGlyGlnProAlaAlaGluSerGlyLysIleAspValGlyAspValI 1543
QY 1601 TCATGGCAATCAGCGCAAGATTGTGCAGACTACACCCCTGGCTGGCTGAGGCTGAGCTGCCCC 1660
Db 1543 leLeuLysValAsnGlyAlaProLeuLysGlyLeuSerGlnGlnAspValIleSerAlaL 1563
QY 1661 TGCAGAAGCGCTGGAATCAGCGCGGCGGCTGATCGACCTGTGTGGTTCGCGCTGCGCCCC 1720
Db 1563 euArg-----GlyThrAlaProGluValSerLeuLeuLeuLeuLeuLeuLeuLeuLeu 1577
QY 1721 CA-----AAGAGTATGACGATGAGCTGACCTCTCTTCTGCTGAAGTCCCAAAA 1765
Db 1577 roAlaProGlyValLeuProGluIleAsp-----ThrThrPheLeu----- 1590
QY 1766 GGGGAAACCAAAATTCACGCGTTAGGAAACAGATGAGCTCGCGCCCCCACCCTCGTGAACACAA 1825
```

```
Db 1591 -----AsnProLeuTyrSerProAlaAsnSerPheLeuAsnSerSerLysGluThrSerG 1609
QY 1826 AGCTCGACGACGCTTCGAGAGAGCCACATGACACACACACACATGGCATCTTGGGACC 1885
Db 1609 InProSerSerValGluGlnGlyAlaSerSerHis-AspAsnGlyValSerGlyLys 1628
QY 1886 TGAATCTATCACCCAGGAATCAAACTCCCTTTGGCCCTGAACAGCCGACAGATGAAGA 1945
Db 1629 ThrLysAsnHis----- 1632
QY 1946 ACAGCTCGGCGACCTTTTGAAGCCCAATGTGGAGGAAAGGGAGCAG----- 1993
Db 1633 ---CysArgAlaProSerArgGluSerTyrSerAspHisSerGluSerGlyGluAsp 1651
QY 1994 -----CCAGCGCTTTGGGAGAAGATCTCAAGGATCCACAGCTCTCATTCCT 2038
Db 1652 AspSerValArgAlaProAlaLysMetProAsnValThrArgValAlaAla----- 1668
QY 2039 TTCTCTCGGCGGAGTAATTTGGTCTCTCCAG-----CTTTGGGG 2080
Db 1669 -----PheProHisGluAlaProArgSerGlnGluSerIleCysAlaMetPheTyr 1686
QY 2081 ACTCTCTCTTGAACCTTAATAAGACCCACTGGAGTCTCTCTCTCATCCCTCTCTCT 2140
Db 1687 LeuProArgLysIleProGlyLys-----LeuGluSerGluSerSerHisProPro 1704
QY 2141 CTG 2143
Db 1705 Leu 1705
RESULT 15
A47747
tight junction protein ZO-1 - human
C:Species: Homo sapiens (man)
C>Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
C:Accession: A47747
R:Willott, E.; Balda, M.S.; Fanning, A.S.; Jameson, B.; Van Itallie, C.; Anderson, J.
Proc. Natl. Acad. Sci. U.S.A. 90, 7834-7838, 1993
A:Title: The tight junction protein ZO-1 is homologous to the Drosophila discs-large
A:Reference number: A47747; MUID:93361541; PMID:8395056
A:Accession: A47747
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1736 <WTL>
A:Cross-references: GB:I14837; NID:q292937; PID:AAA02891.1; PID:q292938
C:Superfamily: guanylate kinase homology; GLGF domain homology
C:Keywords: alternative splicing; membrane protein; phosphoprotein
F:15-94/Domain: GLGF domain homology <GLGF>
F:181-248/Domain: GLGF domain homology <GLGF2>
F:416-486/Domain: GLGF domain homology <GLGF3>
F:633-782/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 5,33e-06 Length: 1736
Score: 220.50 Matches: 135
Percent Similarity: 36.71% Conservative: 108
Best Local Similarity: 20.39% Mismatches: 244
Query Match: 5.35% Indels: 175
DB: 2 Gaps: 28
US-09-502-945-4 (1-2236) x A47747 (1-1736)
```

```
QY 358 GTGCGTCTGACCGCTGTGCACCCCGGAGGCGCTCGAGTGTGGTGGCTGGAG 417
Db 12 ValThrLeuHisArgAlaProGlyPheGlyPheGlyIleAlaIleSerGlyGlyArgAsp 31
QY 418 -----TTTGGCTGTGG-----CTCTCATCTCCACCTCATCAAGCGGT 459
Db 32 AsnProHisPheGlnSerGlyGluThrSerIleValIleSerAspValLeuLysGlyGly 51
QY 460 CAGGCAGACAGCGTGGGCTCCAGGATGAGGAGGAGATCGTCGGATCATGATATCC 519
```

```

Db 52 ProAlaGluGly- ---GlnLeuGlnGluAsnAspArgValAlaMetValAsnGlyValSer 70
QY 520 ATCTCTCTCTGACCCATGAGGAGTCAACACCTCATTCGA ---ACCAAGAAACTGTG 576
Db 71 MetAspAsnValGluHisAlaPheAlaValGlnGlnLeuArgLysSerGlyLysAsnAla 90
QY 577 TCCATCAAGTGGAGACACATCGGC ---CTGATCCCGCGTAAAGCTCTCTCTGATGAG 630
Db 91 LysIleThrIleArgArgLysLysValGlnIleProValSerArgProAspProGlu 110
QY 631 CCCCTCACTTGGCAGTATGTGATCAGTTTGTGTGCGGAATCTGGGGGTGGAGGCAGC 690
Db 111 ProValSerAspAsnGluAspSerTyrAspGluGluIleHisAspProArgSerGly 130
QY 691 CTGGGCTCCCTGGAAATCGGGAACCAAGAG --- 723
Db 131 ArgSerGlyValValAsnArgSerGluLysIleTrpProArgAspArgSerAlaSer 150
QY 723 --- 723
Db 151 ArgGluArgSerLeuSerProArgSerAspArgArgSerValAlaSerSerGlnProAla 170
QY 724 AAGAGGTCTTCATCAGCCTGGTAGGCTCCCGA ---GGCCTTGGCTGCAGCAT 774
Db 171 LysProThrLysValThrLeuValLysSerArgLysAsnGluGluTyrGlyLeuArgLeu 190
QY 775 TCCACGGCCCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAACCTGGCTCCCTG 834
Db 191 AlaSer ---HisIlePheValLysGluIleSerGlnAspSerLeu 204
QY 835 TCTGTCTAGGTGGGA ---TTGGAGATAGGGACACAGATTGCGAAGTCAATGGCTCGAC 891
Db 205 AlaAlaIaArgAspGlyAsnIleGlnGluGlyAspValValLeuLysIleAsnGlyThrVal 224
QY 892 TTCTCTAACCTGGATCAACAAGAGGCTGTAAATGTGCTGAAATAAGCGCAGCCTGACC 951
Db 225 ThrGluAsnMetSerLeuThrAspAlaLysThrLeuIleGluArgSerLysGlyLysLeu 244
QY 952 ATCTCCATTTAGCTGACGTGGCGGAGCTGTTTC ---ATGACAGAC 996
Db 245 LysMetValValGlnArgAspGluArgAlaThrLeuLeuAsnValProAspLeuSerAsp 264
QY 997 CGGGAGCGCTGGCAGAGCGCGGAGCGT ---GAGCTCAGCGCGCAGAG 1044
Db 265 SerIleHisSerAlaAsnAlaSerGluArgAspIleSerGluIleGln --- 281
QY 1045 CTCTCATGACAGCGCTGGCATGGATGAGTCCAAAGATCTCTCCAGGAGCAGCAGAG 1104
Db 282 ---SerLeuAlaSerAspHisSerGlyArgSerHisAspArg 294
QY 1105 ATGGAGCGGCAAGAGAGAAAGAAATGCCAGAGAGCAGGAGGAAATGAGAGATAC 1164
Db 295 ProProArgArgSerArgSerArgProAspGlnArgSerGluProSerAspHisSer 314
QY 1165 CGGAGGAGATGGACAG ---ATGTAGAGGAGAGACAGAAAGTTT 1206
Db 315 ArgHisSerProGlnGlnProSerAsnGlySerLeuArgSerArgAspGluArgIle 334
QY 1207 AAGAAGCAATGGGAAGAAGACTGGGGCTCAAAGGAACAGCTACTCTTCCTTAACACCATC 1266
Db 335 SerLys ---ProGlyAlaVal 340
QY 1267 ACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGATGATGATCAGGAGTGGAACT 1326
Db 341 SerThrProValLysHisAlaAspAspHisThrProLysThrValGluGluVal --- 358
QY 1327 GAGCTCGAGCCCGCAGATGACCTGGATGGAGCAGCAGGAGGAGAGCAGGATTC 1386
Db 359 ---ThrValGluArgAsnGluLysGlnThr 367
QY 1387 CGGAAATATGAGAGGCTTTTGACCCCTACTATGTTTCACCCAGCAGCAG --- 1437
Db 368 ProSerLeuProGluProLysProValTyrAlaGlnValGlyAsnGlnMetTrpIleTyr 387

```

```

QY 1438 ---ATCATGGG 1446
Db 388 LeuSerValHisLeuMetValSerTyrLeuIleGlnLeuMetLysMetGlyPheLeuArg 407
QY 1447 AAGATGTCGGCTCCCTACGATCAAGAGGAGGATCTCTAGACCTGGCCCTGGAAGGC 1506
Db 408 ProSerMetLysLeuValLysPheArgLysGlyAspSerValGlyLeuArgLeuAlaGly 427
QY 1507 GGTGTGACTCCCATTTGGGAGGTGCTGTTCTGTGTGTATGACGGGGGAGCTGCT 1566
Db 428 GlyAsnAsp ---ValGly ---IlePheValAlaGlyValLeuGluAspSerProAla 444
QY 1567 GAGCGCATGTGTGATGTGAAGGAGGACGAGATCATGCGCAATCAACGGCAAGATTGTG 1626
Db 445 AlaLysGlu ---GlyLeuGluGluGlyAspGlnIleLeuArgValAsnValAspPhe 463
QY 1627 ACAGACTACACCTGGCTGAGGCTGACGCTCCCTGCAGAGGCTGGAATCAGGCGGG 1686
Db 464 ThrAsnIleIleArgGluGluAlaValLeuPheLeu --- 475
QY 1687 GACTGGATGACACCTTGTGTGCTGCGCTGCCCCCAAGGAGTATGACGATGAGCTGACC 1746
Db 476 ---LeuAspLeu ---ProLys ---GlyGluGluValThr 485
QY 1747 TCTTGTGTGAAGTCCAAAGGGGAAACCAATTCACGCGTTAGGAAACAGTGCCTCCG 1806
Db 486 IleLeuAlaGlnLysLysLys ---AspValTyr 495
QY 1807 CCCACCTCTGACACACAAAGCCCTCGGACAGCCCTTCAG ---AGAGCCACATGACACAC 1863
Db 496 ArgArgIleValGluSerAspValGlyAspSerPheTyrIleArgThrHisPheGluTyr 515
QY 1864 ACCAGATGGCATCCTTGGGACCTGAATCTATCACCCAGGAATCTCAAACTCCCTTTGGCC 1923
Db 516 GluLysGluSerProTyrGly ---LeuSer 524
QY 1924 CTGAACAGGCGCCAGATAAGAAACAGCTCGGGGCCACTTTTGAAGCCCAATGTGGAGGA 1983
Db 525 PheAsnLysGlyGluValPheArgAlaValAspThrLeuTyrAsnGlyLysLeu ---Gly 543
QY 1984 AAGGAGCAGCCACCCCTTTGGGAGAAGATCTCAAGGATCCAGACTCTCATTCTTCTTCT 2043
Db 544 SerTrpLeuAlaIleArgIleGlyLysAsnHisLysGluValGluArgGlyIleIlePro 563
QY 2044 ---CTGGCCAGTGAATTTGGTCTCTCCCA ---GCTTTG 2076
Db 564 AsnLysAsnArgAlaGluGlnLeuAlaSerValGlnTyrThrLeuProLysThrAlaGly 583
QY 2077 GGGGAC 2082
Db 584 GlyAsp 585

```

Search completed: March 21, 2003, 13:05:51
Job time : 77.8301 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	2823	68.5	552	1	A175_HUMAN	Q9Y6n9 homo sapien
2	275.5	6.7	933	1	Z03_HUMAN	Q95049 homo sapien
3	268.5	6.5	905	1	Z03_MOUSE	Q9QXV1 mus musculu
4	263.5	6.4	767	1	D1G4_HUMAN	P78352 homo sapien
5	261.5	6.3	724	1	D1G4_RAT	P31016 rattus norv
6	261.5	6.3	1745	1	Z01_MOUSE	P39447 mus musculu
7	256.5	6.3	724	1	D1G4_MOUSE	Q62108 mus musculu
8	249.5	6.1	898	1	Z03_CANFA	O62683 canis fami
9	244.5	5.9	904	1	D1G1_HUMAN	Q12959 homo sapien
10	243	5.9	849	1	D1G3_MOUSE	P70175 mus musculu
11	243	5.9	849	1	D1G3_RAT	Q62936 rattus norv
12	240.5	5.8	911	1	D1G1_RAT	Q62968 rattus norv
13	234.5	5.7	817	1	D1G3_HUMAN	Q92796 homo sapien
14	230	5.6	870	1	D1G2_HUMAN	Q15700 homo sapien
15	229	5.6	852	1	D1G2_RAT	Q63622 rattus norv
16	220.5	5.4	1736	1	Z01_HUMAN	Q07157 homo sapien
17	195.5	4.7	2485	1	PTND_HUMAN	Q12923 homo sapien
18	193.5	4.7	2161	1	SHK1_HUMAN	O9V566 homo sapien

CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AF039700; AAC18049.1; ALT_FRAME.
 DR EMBL; AF039699; AAC18048.1; -
 DR EMBL; AB006955; BAA81739.1; -
 DR EMBL; AB018687; BAA81740.1; -
 DR HSSP; P29476; IOAV.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 6.
 DR SMART; SM00228; PDZ; 3.
 DR PROSITE; PS50106; PDZ; 3.
 KW Antigen; Repeat; Alternative splicing; Polymorphism.
 FT DOMAIN 87 169 PDZ 1.
 FT DOMAIN 211 293 PDZ 2.
 FT DOMAIN 452 537 PDZ 3.
 FT VARSPLIT 1 31 MISSING (IN ISOFORM 2).
 FT VARSPLIT 404 552 MISSING (IN ISOFORM 3).
 FT VARIANT 519 519 D -> E (IN DBSNP:1064074).
 /FTID=VAR_012320.
 FT CONFLICT 103 103 R -> S (IN REF. 2; BAA81739).
 FT CONFLICT 280 280 N -> S (IN REF. 2).
 FT CONFLICT 305 305 A -> T (IN REF. 2; BAA81739).
 SQ SEQUENCE 552 AA; 62223 MW; 7B405B02CACE8241 CRC64;

Alignment Scores:

Pred. No.: 1.15e-145 Length: 552
 Score: 2823.00 Matches: 551
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 68.52% Indels: 0
 DB: 1 Gaps: 0

US-09-502-945-4 (1-2236) x A175_HUMAN (1-552)

QY 97 ATGGACCGAAAGTGGCCGAGAAATTCGGCATAGGTGGATTTCTGATTGAAATGAT 156
 DB 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
 QY 157 GCAGAGAAGGACTATCTCTATGATGTCTGCGAATGTACCCAGACCATGGACGTGCC 216
 DB 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
 QY 217 GTGCTGTGGGAGACCTGAAGCTGTGCATCAATGAACCCAGCCGCTGCCTCTGTTGAT 276
 DB 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
 QY 277 GCATTGCGCCGCTGATCCACTGAAGCACCAGGTGGATATGATCAGTACCCCGCGG 336
 DB 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 QY 337 CGCTCCAGGAAGCTGAAGAGGTGGCTCTGGACCGCTCTGCACCCCGAAGCGCTCGCCCTG 396
 DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
 QY 397 AGTGTGCTGGTGGCTGGAGTTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGGC 456
 DB 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
 QY 457 GGTCAAGCAGACAGCTGGCGCTCCAGGTAGGAGGAGATCGTCCGGATCAATGATAT 516
 DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 QY 517 TCCATCTCTCTGTACCCATCAGGAGGTCAATCAACTCATTCGAACCAAGAAACTGTG 576
 DB 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
 QY 577 TCCATCAAAAGTGAGACACATCGCCCTGTATCCCGGTGAAAGAGCTCTCCTGATGAGCCCTC 636

DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerProAspGluProLeu 180
 QY 637 ACTTGGCAGTATCGATCAGTTGTGTGGAAATCTGGGGCGTGGAGCAGCAGCTGGGC 636
 DB 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlySerLeuGly 200
 QY 697 TCCCTCGAAATCGGAAACAAGAGAGAGAGTCTTCATCAGCCTGTGTAGCTCCCGA 756
 DB 201 SerProGlyAsnArgGluAsnLysGluLysLysValPheIleSerLeuValGlySerArg 220
 QY 757 GGCTTGGCTGCAGCATTTCCAGCGCCCATCCAGAAGCCTGSCATCTTTATCAGCAT 816
 DB 221 GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
 QY 817 GTGAACCTCGCTCCCTGTCTGTAGGTGGATTTGAGATAGGGACCAAGATTTGCGAA 876
 DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
 QY 877 GTCAATGGCTGCAGTCTCTTAACCTGGATCACAAGAGGCTGTAAATGTGCTGAAAT 936
 DB 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
 QY 937 AGCGCAGCTGCACCATCTCCATTGTAGTGCAGTGGCGGGAGCTGTTTCATGACAGAC 996
 DB 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
 QY 997 CGGAGCGCTGGCAGAGCGCGCAGCTGAGCTGCAGCGGAGGAGCTTCTCATGCAG 1056
 DB 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
 QY 1057 AAGCGCTGGGATGGAGTCCACAAGATCTCTCAGAGCAGCAGGAGATGGAGCGCAA 1116
 DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnMetGluArgGln 340
 QY 1117 AGGAGAAAGAAATGCCAGAGCAGCAGAGAAATCAGAGATACCGGAAGGAGATG 1176
 DB 341 ArgArgLysGluIleAlaGlnLysAlaGluLysGlnArgTyrArgLysGluMet 360
 QY 1177 GAACAGATTGTAGAGGAGAGAGAGTAAAGCAATGGGAAGAGACTGGGGCTCA 1236
 DB 361 GluGlnIleValGluGluGluLysLysLysGlnTrpGluGluAspTrpGlySer 380
 QY 1237 AAGGAACAGCTACTTTCCTAAACCATCACTCTCAGGTACACCCAGTACCCCTTCGC 1296
 DB 381 LysGluGlnLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 QY 1297 AAGCAAAGTATGATCAGGAGTGAACCTGAGCTCCAGCCCGCAGATGACCTGGATGA 1356
 DB 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspAspLeuAspGly 420
 QY 1357 GGCACGGAGGAGCAGGAGCAGGATTCGGAAATATGAGGAAGCTTTGACCCCTAC 1416
 DB 421 GlyThrGluGluGlnGlyGluAspPheArgLysTyrGluGluGlyPheAspProTyr 440
 QY 1417 TCTATGTTACCCAGACAGACATCATGGGAAGATGTCCGGCTCCTACGCATCAAGAAG 1476
 DB 441 SerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLys 460
 QY 1477 GAGGATCTTAGACCTGGCCCTGGAAGCGGTGTGACTCCCTCCCATTTGGGAAGTGGTC 1536
 DB 461 GluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProIleGlyLysValVal 480
 QY 1537 GTTCTGTGTGTATGACGGGGAGCTGCTGAGCGGATGGTGGCATTTGTGAAAGGGAC 1596
 DB 481 ValSerAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAsp 500
 QY 1597 GAGATCATGCAATCAACGGCAAGATTGTACAGACTACACCTCGCTGAGCTGAGCTGAGCT 1656
 DB 501 GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAla 520
 QY 1657 CCCTGCAAGAGGCTTGAATCAGGGGGGAGCTGGATCGACCTGTGGTGGCTGCTGC 1716

```

Db 521 AlaLeuGlnLysAlaTrpAsnGlnGlyGlyAspTrpIleAspLeuValAlaValCys 540
QY 1717 CCCCCAACGAGTATGACGATGAGTGCCTTC 1749
Db 541 ProProLysGluTyrAspAspGluLeuThrPhe 551
RESULT 2
Z03_HUMAN STANDARD; PRT; 933 AA.
AC 095049;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
DE occludens 3 protein) (Tight junction protein 3).
GN TJP3 OR Z03.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Brower A., Gordon L., Dias J., Ramirez M.,
RA Stillwagen S., Phan H., Velasco N., Do L., Regala W., Terry A.,
RA Ganes J., Danganan L., Erier A., Christensen M., Georgescu A.,
RA Avila J., Liu S., Attix C., Andreise T., Trankheim M.,
RA Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R.,
RA Thomas P., Quan G., Krommiller B., Arellano A., Sanders C., Ow D.,
RA Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.,
RT "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
RT serine protease gene cluster."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: INTERACTS WITH OCLUDIN, CLAUDINS AND ZO-1 (BY
CC SIMILARITY).
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AC005954; AAC72274.1; ALT_INIT.
CC HSP; P31016; 1BFE.
CC Genew; HGNC:11829; TJP3.
CC InterPro; IPR000619; Guanylate_kin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00595; PDZ; 3.
CC Pfam; PF00625; Guanylate_kin; 1.
CC SMART; SM00072; GuK; 1.
CC SMART; SM00228; PDZ; 3.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
CC PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
CC PROSITE; PS50002; SH3; 1.
CC PROSITE; PS50106; PDZ; 3.
CC Tight junction; SH3 domain; Repeat; Membrane.
KW DOMAIN 11 93
FT DOMAIN 195 272 PDZ 1.
FT DOMAIN 394 460 PDZ 2.
FT DOMAIN 489 563 PDZ 3.
FT DOMAIN 675 775 GUANYLATE_KINASE.
SQ SEQUENCE 933 AA; 102800 MW; 680298CFD0615B47 CRC64;
Alignment Scores: 8.16e-08 Length: 933
Pred. No.:

```

```

Score: 275.50 Matches: 142
Percent Similarity: 35.97% Conservative: 90
Best Local Similarity: 22.02% Mismatches: 203
Query Match: 6.69% Indels: 203
DB: 1 Gaps: 28

US-09-502-945-4 (1-2236) x Z03_HUMAN (1-933)
QY 385 GGCTCTGGCGCTGAGTGTGGCTGGCTGGAGTTT---GGCTGTGGCTCTTCATCTCC 441
Db 21 GlyPheGlyIleAlaIleSerGlyArgAspArgProGlyGlySerMetValValSer 40
QY 442 CACCTCATCAAAAGCGGTGACGACAGACAGCGTCGGGCTCCAGGTAGGAGGACGATCTGC 501
Db 41 AspValValProGlyGlyProAlaGluGly---ArgLeuGlnThrGlyAspHisIleVal 59
QY 502 CGGATCAATGATATTCATCTCCTCTCTGATCCCATGAGGAGGATCATCAACATCTATTCGA 561
Db 60 MetValAsnGlyValSerMetGluAsnAlaThrSerAlaPheAlaIleGlnIleLeuLys 79
QY 562 ACC---AAGAAACTGTGTCTCATCAAAAGTG-----AGACACATCGGCTGATCCCC 609
Db 80 ThrCysThrLysMetAlaAsnIleThrValLysArgProArgArgIleHisLeuProAla 99
QY 610 GTGAAAGCTCTCTCT---GATGAGCCCTC 636
Db 100 ThrLysAlaSerProSerSerProGlyArgGlnAspSerAspGluAspGlyProGln 119
QY 637 ACTGGCAGTAGTGTGGATCAG-----TTTGTCGCGAATCTGGG 675
Db 120 ArgValGluGluValAspGlnGlyArgGlyTyrAspGlyAspSerSerGlySerGly 139
QY 676 -----GSCGTGCGAGGAGCGCTGGGCTCC 699
Db 140 ArgSerTrpAspGluArgSerArgArgProArgProGlyArgArgGlyArgAlaGlySer 159
QY 700 CTGGAATCGGAAACAAGAGAGAGAGGTCTTCATCGCTGGTGGTATCGCGAGGC 759
Db 160 HisGlyArgArgSerProGlyGly-----GlySerGluAla 171
QY 760 CTGCTGCATCAGTTCCTCAGCGGC-----CCCATCCAG-----AAG 795
Db 172 AsnGlyLeuAlaLeuValSerGlyPheLysArgLeuProArgGlnAspValGlnMetLys 191
QY 796 CCT----- 798
Db 192 ProValLysSerValLeuValLysArgArgAspSerGluGluPheGlyValLysLeuGly 211
QY 799 ---GGCATCTTTATCAGCCATGTGAAACCTGGCTCCCTGTCTGCT---GAGGTGGGATTG 852
Db 212 SerGlnIlePheIleLysHisIleThrAspSerGlyLeuAlaIaArgHisArgGlyLeu 231
QY 853 GAGATAGGCGACGATGTCGAGTCAATGGCTGCGACTTCTTAACCTCGATCACAAG 912
Db 232 GlnGluGlyAspLeuIleGlnGlyValSerSerGlnAsnLeuSerLeuAsn 251
QY 913 GAGCGTGTAAATGTCTGAAAAATAGCCGACCTGACCATCTCCATGTAGTACGACT 972
Db 252 AspThrArgArgLeuIleGluLysSerGluGly---LysLeuSerLeuValLeuArg 270
QY 973 GCGCGGAGCTGTCATGACAGCGGGAGCGGTGGCAGAGCGCGCGGCGGCGGAGCTG 1032
Db 271 AspArgGlyGlnPheLeuValAsnIleProProAlaValSerAspSerAspSerPro 290
QY 1033 CAGCGGACGAGGCTTCTCATCAGAGCGGTGGCGATGAGTCCACACAGATCCTCCAG 1092
Db 291 LeuGluGluGlyValThrMetAlaAspGluMetSerSerProProAlaAspIleSerAsp 310
QY 1093 GAGCAGCAGGAGTAGGACGG----- 1113
Db 311 LeuAlaSerGluLeuSerGlnAlaProProSerHisIleProProProProArgHisAla 330
QY 1114 CAAGGAGAAAAGAAATTTGCCAG---AAGCAGCAGAGGAAAATGAGAGATACCGGAAG 1170

```



```

Db 60 MetValAsnGlyValSerValGluAsnValThrSerAlaPheAlaIleGlnIleLeuLys 79
QY 562 ACC---AGAAAACGTGTCCATCAAGTG-----AGACATCGCGCTGATCCCC 609
Db 80 ThrCysThrLysThrAlaAsnValThrValLysArgProArgArgValGlnLeuProAla 99
QY 610 GTGAAGACTCTCCCGATGATGAGCCCTCACTTGGCAGTATGTGGATCGATTGTCGGA 669
Db 100 ThrLysAlaSerProAlaSer-----GlyHisGlnLeuSerAspGlnGluAlaAsp 117
QY 670 TCTGGG----- 675
Db 118 HisGlyArgGlyTyrGluGlyAspSerSerGlySerGlyArgSerTrpClyGluArg 137
QY 676 -----GGCGTGGAGCAGCCCTGGCTCCCTCGAATCGGAAAC 717
Db 138 SerArgArgSerArgAlaGlyArgArgGlyValGlySerHisGlyArgArgSerSer 157
QY 718 -----AAG 720
Db 158 GlyGlyGlySerGluAlaAsnGlyLeuAspLeuValSerGlyTyrLysArgLeuProLys 177
QY 721 GAG-----AAGAGGTCTTCATCAGCCTGGTAGGCTCCCGAGGC 759
Db 178 GlnAspValLeuMetArgProLeuLysSerValLeuValLysArgArgAsnSerGluGlu 197
QY 760 CTGGCTGCAGCATTCAGCGGCCCTCCAGAGCCTGGATCTTTATCAGCATGTG 819
Db 198 PheGlyValLysLeuGlySer-----GlnIlePheIleLysHisIle 211
QY 820 AAACCTGGCTCCTCTCTGCT---GAGGTGGATTGGAGATGAGATGAGGCTGTAATGTGCTGAA 936
Db 232 IleAsnGlyValSerSerAlaAsnLeuSerLeuSerAspThrArgArgLeuIleGluLys 251
QY 937 AGCCGC---AGCCTACCATCTCCATTGTAGCTGCAGCTGGCGGGAGCTG----- 984
Db 252 SerGluGlyGluLeuThrLeuLeuValLeuArgAspSerGlyGlnPheLeuValAsnIle 271
QY 985 -----TTCATGCAGAC-----CGGAG 1002
Db 272 ProProAlaValSerAspSerAspSerSerLeuMetGluAspIleSerAspLeuThrSer 291
QY 1003 CGGTGGCAGAGCGCGCAGCGTGCAGCTGCAGCGCAGGAGCTTCTCATCAGAAGCGG 1062
Db 292 GluLeuSerGlnAlaProProSerHisValProProProLeuLysGlyGln---Arg 310
QY 1063 CTGGCGATGGAGTCCAAACAGATCTCCAGGAGCAGCAGGAGATGGAGCGGCAAGGAGA 1122
Db 311 SerProGluAspSerGlnThrAspSerProValGluThrProGlnProArgArgGlu 330
QY 1123 AAAGAATATGCCAGAGCGCAGCAGAGAAATGAGACATACCCGGAAGAGATGGAACAG 1182
Db 331 ArgSerValAsnSerArgAlaIleAlaGluProGluSer----- 343
QY 1193 ATTGTAGAGGAGGAAGAGAGTTTAAGAAGCAATGGGAAGAGACTGGGGCTCAAGAA 1242
Db 343 ----- 343
QY 1243 CAGCTACTCTTGCTTAAACCATCATCTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCA 1302
Db 344 -----ProGlyGluSerArgTyrAspIleTyrArgValProSerArgGlnSer 359
QY 1303 AAGTATGATCAGGAGTGGAACTCAGCTCCGAGCCGAGATGACCTGGATGGAGGCAGC 1362
Db 360 LeuGluAspArgGlyTyrPro----- 367
QY 1363 GAGGAGCAGGAGAGCAGGATTTCCGGAATATGAGGAAGGCTTTGACCCCTACTCTATG 1422
Db 367 ----- 367

```

```

QY 1423 TTCACCCAGACGATCATGGGAAGGATGTCGGGCTCTCATGACATCAAGAAGGAGGA 1482
Db 368 -----AspThrArgValValSerPheProLysGlyAla 378
QY 1493 TCCTTAGACCTGGCCCTGGGAAGGGGTGTGGACTCCCCCATTTGGGAAGGTGGTCTTCT 1542
Db 379 SerIleGlyLeuArgLeuAlaGlyGlyAsnAsp-----ValGly---IlePheValSer 395
QY 1543 GCTGTGTATGAGCGGGGAGCTGCTGAGCGCATGGTGCATGTGAAAGGGGACGAGATC 1602
Db 396 GlyVal---GlnAlaGlySerProAlaAspGlyGlnGlnGlnGluGlyAspGluIle 414
QY 1603 ATGCAATCAACGCAAGATTGTGACAGACTACACCTGGCTGGCTGAGGCTGACGCTGCCCTG 1662
Db 415 LeuGlnValAsnGlyMetProPheArgAsnLeuThrArgGluAla----- 430
QY 1663 CAGAAGCCTGGATCAGGCGGGAGCTGGATCGACCTTGTGGTTCGCTGCGCTGCCGCCCA 1722
Db 431 -----ValGlnPheLeuLeuGlyLeu---ProPro 439
QY 1723 AAGGATGATGACGATGAGCTGACCTTCTGTCTGAAGTCCAAAGGCGGAACCAATTCAC 1782
Db 440 GlyGlu-----AspMetGluLeuValThrIleSerLysThrGlyHisSerLeuArg 456
QY 1783 GCGTTAGAAACAGTGTGCTCCGCCCCACCTCGTGAACACAAAGCCTCGGACGACGCTT 1842
Db 457 ArgTrpSerGlnSerArgVal----- 463
QY 1843 GAGAGAGCCACATGACACACACAGATGGCATCTCTTGGGACCTGAATCATCACCCAGG 1902
Db 464 -----GlyAspSerPheTyrIleArgThrHis---PheGluLeuGluProSerPro--- 479
QY 1903 AATCTAAACTCCCTTGGCCCTGAACGAGGCGCAGATA----- 1941
Db 480 -----ProTyrGlyLeuGlyPheThrArgGlyAspValPheHisValValAspThrLeu 497
QY 1942 -----AGGAACAGCTCGGCCACTTTTGAAGCCCAATGTGGAGGAAGGAGGAGCACC 1995
Db 498 TyrProGlySerGlyProGlyHisGlyHisSerHis---GlyGlyLeuTrpLeuAla 516
QY 1996 AGCGCTTTGGAGAAGATCTCAAGATCCAGACTCTCATCTCTCTTCT 2043
Db 517 AlaArgMetGlyArgAspLeuArgGlnGluArgGlyValIlePro 532

```

RESULT 4

DLG4_HUMAN STANDARD; PRT; 767 AA.

AC P78352; Q92941; Q9UKK8;

DT 01-NOV-1997 (Rel. 35, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Presynaptic density protein 95 (PSD-95) (Discs, large homolog 4)

DE (Postsynaptic density-95).

GN DLG4 OR PSD95.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RX MEDLINE=9743282; PubMed=9286702;

RA Stathakis D.G., Hoover K.B., You Z., Bryant P.J.;

RT "Human postsynaptic density-95 (PSD95): location of the gene (DLG4)

RT and possible function in nonneural as well as in neural tissues.";

RL Genomics 44:71-82(1997).

RN [2]

RP REVISIONS.

RC TISSUE=Mammary gland;

RA Stathakis D.G., Hoover K.H., You Z., Bryant P.J.;

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

Pred. No.:	4,47e-07	Length:	724
Score:	261.50	Matches:	125
Percent Similarity:	40.45%	Conservative:	93
Best Local Similarity:	23.19%	Mismatches:	206
Query Match:	6.35%	Indels:	115
DB:	1	Gaps:	24

US-09-502-945-4 (1-2236) x DLG4_RAT (1-724)

[illegible][illegible]

RESULT 6

Z01_MOUSE	
ID	Z01 MOUSE
PRT:	1745 AA.

AC P39447;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Tight junction protein ZO-1 (zonula occludens

DE occludens 1 protein) (Tight junction protein 1).

GN TJPL OR ZO1.

Mus musculus (Mouse)

OC Eukaryota; Metazoa;

OC MAMMALIA; EUTHERIA;
OV NEST MAR-10000.

```

    NCBL_TaxID=10090;
    PN [1]

```

RN {T}
RD SEQUENCE FROM N A

RF SEQUENCE FROM N.A.
PC STPATN=129.

RC DUKHAN=122,
RX MEDLINE=93252986. DUKH

RA MEDLINE 55252500, 1984
RA Ttoh M. Nagafuchi A

RA Tsukita S.

The 220-kD protein

RT cells is identical to

RT epithelial cells: cDNAs

R. J. Cell Biol. 121:491

CC - I - FUNCTION: THE N-T

CC REQUIRED FOR TIGHTENING THE ...

CC HAVE SPECIFIC PRO

MIGHT BE INVOLVED

- I - SUBUNIT: INTERACT

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

CC -|- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE, CYTOPLASMIC SIDE.
 CC MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT
 CC OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT (BY SIMILARITY).
 CC -|- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -|- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
 CC -|- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: D14340; BAA03274.1; -
 DR PIR: A46431; A46431.
 DR HSP: P31016; LBE9.
 DR MGD: MGI:98759; Tjp1.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR000906; ZUS.
 DR Pfam: PF00018; SH3; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00791; ZUS; 1.
 DR SMART: SM00072; GuK; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR SMART: SM00218; ZUS; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
 DR PROSITE: PS00502; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS0106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 DR Tight junction; SH3 domain; Repeat; Membrane.
 FT DOMAIN 23 110
 FT DOMAIN 186 264
 FT DOMAIN 421 502
 FT PDZ 3.
 FT SH3.
 FT DOMAIN 516 584
 FT DOMAIN 644 794
 FT GUANYLATE KINASE.
 FT DOMAIN 1242 1247
 FT POLY-PRO.
 FT DOMAIN 1424 1430
 FT POLY-PRO.
 SQ SEQUENCE 1745 AA; 194710 MW; C3DA2C0A9F411F66 CRC64;

Alignment Scores:
 Pred. No.: 5,17e-07 Length: 1745
 Score: 261.50 Matches: 143
 Percent Similarity: 36.73% Conservative: 106
 Best Local Similarity: 21.09% Mismatches: 222
 Query Match: 6.35% Indels: 207
 DB: 1 Gaps: 30

US-09-502-945-4 (1-2236) x ZOL_MOUSE (1-1745)

QY 358 GFGCCTGGACGCTGTCACCCGAGCCCTCGCCTGAGTGCGGTGGCTGGAG 417
 Db 24 ValThrLeuHisArgAlaProGlyPheGlyPheGlyAlaHisSerGlyGlyArgAsp 43
 QY 418 -----TTTGGCTGGG-----CTCTTCATCTCCACCTCATCAAGCGCGT 459
 Db 44 AsnProHisPheGlnSerGlyGluThrSerLeuValIleSerAspValLeuLysGlyGly 63
 QY 460 CAGGCAGACGCGCTCCAGGTAGGACGAGATCGTCCGATCAATGGATATCC 519
 Db 64 ProAlaGluGly---GlnLeuGlnGluAsnAspArgValAlaMetValAsnGlyValSer 82
 QY 520 ATCTCTCTCTGACCATGAGGAGGAGTGCATCAAC----- 552
 Db 83 MetAspAsnValGluHisAlaPheAlaValGlnGlnLeuArgLysSerGlyLysAsnAla 102
 QY 553 -----CTCATTCGAACCAAGAAACTGTGTCCATCAAGTGAGACACATCGGCTGATC 606

Db 103 LysIleThrIleArgArgLysLysValGlnIleProValSerHis----- 118
 QY 607 CCCGTGAAAGCTCTCTCTGAT---GAGCCCTCCTACCTGGCAGTATGTGATCAGTTTGTG 663
 Db 119 -----ProAspProGluProValSerAspAsnGluAspAspSerTyrAsp 133
 QY 664 TCGGAATCTGGGGCTGCGAGCGCTGGCTCCCTCGAAATCGGGAACCGGAAACAAAGAG 723
 Db 134 GluGluValHisAspProArgAlaGlyArgGlyAlaLeuAlaAsnArgArgSerGluLys 153
 QY 723 ----- 723
 Db 154 SerTrpAlaArgAspArgSerAlaSerArgGluArgSerLeuSerProArgSerAspArg 173
 QY 724 -----AAGAAAGCTCTTCATCAGCTAGCTAGGTCCCGA 756
 Db 174 ArgSerValAlaSerSerGlnProAlaLysProThrLysValThrLeuValLysSerArg 193
 QY 757 -----GGCCTTGGCTGCAGCATTTCCAGCGCCCTCCAGAACCTCCG 801
 Db 194 LysAsnGluGluTyrGlyLeu-----ArgProAla 203
 QY 802 -----ATCTTTATCAGCCATGTGAACCTGGCTCCCTGTCTGTCTGAGGTGGGA---TTG 852
 Db 204 SerHisIlePheValLysGluIleSerGlnAspSerLeuAlaAlaArgAspGlyAspIle 223
 QY 853 GAGATAGGGGACACAGATTGTGCAAGTCAATGGCTGCGACTTCTTAACCTGGATCACAAG 912
 Db 224 GlnGluGlyAspValValLeuLysIleAsnGlyThrValThrGluAsnMetSerLeuThr 243
 QY 913 GAGCTCTAAATGTCTGAAATAATAGCCGACCTGACCATCTCCATTGTAGTGCAGCT 972
 Db 244 AspAlaLysThrLeuIleGluArgSerLysGlyLysLeuLysMetValValGlnArgAsp 263
 QY 973 GCCCGGAGCTGTCATG----- 990
 Db 264 GluArgAlaThrLeuLeuAsnValProAspLeuSerAspSerIleHisSerAlaAsnAla 283
 QY 991 ACAGACGGGAGCGCTGGCAGAGCGCGCGAG-----CGTGAG 1029
 Db 284 SerGluArgAspAspIleSerGluIleGlnSerLeuAlaSerHisSerGlyArgSer 303
 QY 1030 CTGACGCGGAGGAGCTTCTCATCGACAGAGCGGCTG----- 1065
 Db 304 HisAspArgProArgArgSerGlnSerArgSerProAspGlnArgSerGluProSer 323
 QY 1066 -----GCCATGGATCCAAAGATCTCTCCAGGACGACGAGGAG 1104
 Db 324 AspHisSerThrGlnSerProGlnGlnProSerAsnGlySerLeuArgSerArgGluGlu 343
 QY 1105 ATGAGCGGCAAGAGAGAAATGCCCAGAGGACGACAGAGGAAATAGAGATAC 1164
 Db 344 GluArgMetSerLysProGlyAlaIleSerThrProValLysHisValAspAspHisPro 363
 QY 1165 CGGAAGAGGATGGACAGAT---GTAGAGGAGAGAGAGAGTTTAAGAAGCAATGGAA 1221
 Db 364 ProLysAlaValGluGluValThrValGluLysAsnGluLys----- 377
 QY 1222 GAAGACTGGGCTCAAGGAAACACTCTTGTGCTAAACCATCATCTGCTGAGGTACAC 1281
 Db 378 -----GlnThrPro 380
 QY 1282 CCAGTACCCCTTCGCAAGCCAAAGTATGATCAGGAGTGGAACTGAGCTCGAG---CCC 1338
 Db 381 ThrLeuProGluProLysProValTyrAlaGlnValGlnProAspValAspLeuPro 400
 QY 1339 GCAGATGACCTGGATGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1398
 Db 401 ValSerProSerAspGlyAlaLeuProAsnSerAlaHisGluAsp----- 415
 QY 1399 GAAGGCTTTGACCCCTACTCTTATGTTTCAACCCAGACGACATCATGTGGGAAGGATGTCGG 1458

```

Db 416 -----GlylleLeuArgProSerMetLys 423
QY 1459 CTCCTACGATCAAGAAGGAGGATCTTACAGCTGGCCCTGGAAGGGGTGTGGACTCC 1518
    |||:||||| |||: ||| ||| ||| ||| |||
Db 424 LeuValLysPheArgLysGlyAspSerValGlyLeuArgLeuAlaGlyLysAsnAsp--- 442
QY 1519 CCCATTGGGAAGGTGCTGCTCTGCTGTATGAGCGGGGAGCTGCTGACGGCGATGGT 1578
    |||:||||| |||: ||| ||| ||| ||| |||
Db 443 ---ValGly---IlePheValAlaGlyValLeuGluAspSerProAlaAlaLysGlu--- 459
QY 1579 GGCATTGTGAAGGGGAGGATGATGCAATCAACGGCAAGATGTCACACAGCTACACC 1638
    |||:||||| |||: ||| ||| ||| ||| |||
Db 460 GlyLeuGluGluGlyAspGlnIleLeuArgValAsnAsnValAspPheThrAsnIleIle 479
QY 1639 CTGGCTGAGGTAGCGCTGCCCTGCAGAGGCTGGAATCAATGAGGGGGAGTGGATCGAC 1698
    |||:||||| |||: ||| ||| ||| ||| |||
Db 480 ArgGluGluAlaValLeuPheLeu-----LeuAsp 489
QY 1699 CTTGTGGTGGCTGCTGCCCTGAGGAGGATGATGAGTACGCTGACCTCTTCTGCTGAAG 1758
    |||:||||| |||: ||| ||| ||| ||| |||
Db 490 Leu-----ProLys-----GlyGluGluValThrIleLeuAlaGln 501
QY 1759 TCCAAAGGGGAACAAATTCACGCTTAGGAACAGTACGCTCCGGCCACCTCGTG 1818
    |||:||||| |||: ||| ||| ||| ||| |||
Db 502 LysLysLys-----AspValTyArgArgIleVal 511
QY 1819 AACCAAGCCTCGGACGAGCCTTGAG---AGAGGCCATGACACACACAGATGGCAT 1875
    |||:||||| |||: ||| ||| ||| ||| |||
Db 512 GluSerAspValGlyAspSerPheTyrlleArgThrHisPheGluTyArgLysLysGluSer 531
QY 1876 CTTTGGGACCTGAATCATCAACCCAGGATCTCAACTCCCTTTGGCCCTGACACAGGCG 1935
    |||:||||| |||: ||| ||| ||| ||| |||
Db 532 ProTyArgLys-----LeuSerPheAsnLysGly 540
QY 1936 CAGATAAGGAACAGCTCGGGCCACTTTTGAAGGCAATGTGGAGGAAGGAGGAGCAGCC 1995
    |||:||||| |||: ||| ||| ||| ||| |||
Db 541 GluValPheArgValValAspThrLeuTyArgAsnGlyLysLeu---GlySerTrpLeuAla 559
QY 1996 ACCGTTTGGGAGAGATCTCAAGGATCAGCTCTCATCTTCCTTCTCTCTCTCTCTCTCT 2043
    |||:||||| |||: ||| ||| ||| ||| |||
Db 560 IleArgIleGlyLysAsnHisLysGluValGluArgGlyIleIleProAsnLysAsnArg 579
QY 2044 -----CTGGCCAGTCAATTTGGTCTCTCCCA-----GCTTGGGGGAC 2082
    |||:||||| |||: ||| ||| ||| ||| |||
Db 580 AlaGluGlnLeuAlaSerValGlnTyThrLeuProLysThrAlaGlyLysAsp 597
RESULT 7
DLG4_MOUSE
ID DLG4_MOUSE STANDARD; PRT; 724 AA.
AC Q62108;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)
DE (Synapse-associated protein 90) (Discs, large homolog 4).
GN DLG4 OR DLG4H OR PSD95.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DBA/2; TISSUE=Brain;
RA Kohmura N., Yagi T.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
CC -!- SUBUNIT: Interacts with DLGAP1/GKAP (By similarity). Is part of a
CC complex with DLGAP1/GKAP, SHANK1 SHANK2 or SHANK3 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Concentrated at synaptic
CC junctions primarily on the presynaptic side. Also found in
CC postsynaptic density of neuronal cells (By similarity).
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.

```

```

CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; D50621; BAA09297.1; -
DR HSP; P31016; IBE9.
DR MGD; MGI:1277959; Dlg4.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
DR PROSITE; PS00002; SH3; 1.
DR SH3 domain; Repeat.
KW DOMAIN 65 151
FT DOMAIN 160 246
FT DOMAIN 313 393
FT DOMAIN 428 498
FT DOMAIN 534 724
FT SEQUENCE 724 AA; 80472 MW; 7EFC99E1FF90BA CRC64;
SQ
Alignment Scores:
Pred. NO.: 6.5e-07 Length: 724
Score: 258.50 Matches: 127
Percent Similarity: 39.74% Conservative: 90
Best Local Similarity: 23.26% Mismatches: 200
Query Match: 6.27% Indels: 129
DB: 1 Gaps: 25
US-09-502-945-4 (1-2236) x DLG4_MOUSE (1-724)
QY 340 TCCAGGAGCTGAAGGAGTGGTCTGACCGCTCTGCACCCGAGGCGCTCGGCTGAGT 399
    |||:||||| |||: ||| ||| ||| ||| |||
Db 155 AlaGluLysIleIleGluIleLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 173
    |||:||||| |||: ||| ||| ||| ||| |||
QY 400 GTGGCTGGTGGC-----CTGGAGTTTGGCTGTGGGCTCTTCTCATCTCCAC 444
    |||:||||| |||: ||| ||| ||| ||| |||
Db 174 IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyValThrLys 193
QY 445 CTCTCAAAAGCGGTGACGAGACAGCGCTCGG---CTCCAGGTAGGGAGGAGATCGTC 501
    |||:||||| |||: ||| ||| ||| ||| |||
Db 194 IleIleGluGlyGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 213
QY 502 CGGATCATGATGATATTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 561
    |||:||||| |||: ||| ||| ||| ||| |||
Db 214 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaLeuLys 233
QY 562 ---ACCAAGAAACTGTGTCTCATCAAGTG-----AGACACATCGGCTGATCCCC 609
    |||:||||| |||: ||| ||| ||| ||| |||
Db 234 AsnThrTyArgValValTyLeuLysValAlaLysProSerAsnAlaTyLeuSerAsp 253
QY 610 GTGAAAGCTCTCTGTAGAGCCCTCCTCCTG---CAGTATGTGGATCAGTTGTGTGTCG 666
    |||:||||| |||: ||| ||| ||| ||| |||
Db 254 SerTyAlaProProAspIleThrThrSerTySerGlnHisLeuAspAsnGluIleSer 273
QY 667 GAATCTGGGGCGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 696
    |||:||||| |||: ||| ||| ||| ||| |||

```

Db 274 HisSer'-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 289
 Qy 697 -----TCCCTCGAAATCGGAA-----AACAAAG 720
 Db 290 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluGluAspIleProArg 309
 Qy 721 GAGAAAGAGTTCATCAGCCTGGTAGCTCCGAGCCCTGGTGGTGCAGCATTTCCAGC 780
 Db 310 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 329
 Qy 781 GGGCCCATCCAGAACGCTGTCATATCATGCCATG-----AACCTGGCTCC 831
 Db 330 GlyGlu---AspGlyGluGlyIlePheHisSerPheIleLeuAlaGlyGlyProAlaAsp 348
 Qy 832 CTGTCTGCTGAGTGGGATGAGATAGGGACCATGTCGAAAGTCAATGGCTGCAC 891
 Db 349 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 366
 Qy 892 TTCCTAACCTGGATCACAAGGAGCTGTAATGTCTGTAAGAAATAGCCGAGCTGACC 951
 Db 367 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 385
 Qy 952 ATCTCCATTGTAGTGCAGCTGGCGGAGCTGTCATGACAGACCGGAGCGCTGGCA 1011
 Db 386 ValThrIleIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
 Qy 1012 GAGCGCGGCGAGCTGAGCTCCAGCGGAGGAGTCTTCATGCAAGACCGCTG----- 1065
 Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
 Qy 1066 -----GGGANGGAGTCCCAAC----- 1080
 Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
 Qy 1080 ----- 1080
 Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheHisPheGlyAsp 458
 Qy 1081 -----AGATCTCCAGGACGACGAGATGGAGCGGCAAGAGGAAAGAAATTTGCC 1134
 Db 459 ValLeuHisValIleAspAlaSerAspGluGluTrpTrpGlnAlaArgValHisSer 478
 Qy 1135 CAGAAGCGCAGCAGAGAAATGATGACATACCGGAGGATGGACAGATGTAGACGAG 1194
 Db 479 AspSerGluThrAsp-----AspIleGlyPheIleProSerLys 491
 Qy 1195 GAAGAGAAGTTTAAGAACCAATGGAA-----GAAGACTGGGCTCA----- 1236
 Db 492 ArgArgValGluArgArgGluTrpSerArgLeuLysAlaLysAspTrpGlySerSerSer 511
 Qy 1237 -----AAGGACACCTACTCTTGCTTAAACCATCACT---GCTGAGGTA 1278
 Db 512 GlySerGlnGlyArgGluAspSerValLeuSerTyrGluThrValThrGlnMetGluVal 531
 Qy 1279 CAC-----CCAGTACCCCTCGCAAGCAAGATGATCATCAGGAGTGGAACTGAG 1329
 Db 532 HisTyrAlaArgProIleIleIleGlyProThrLysAspArgAlaAsnAspAspLeu 551
 Qy 1330 CTCGAGCCCGCAGATGACCTTGATGGAGCGCAGGAGGAGGAGGAGGATTTCCGG 1389
 Db 552 LeuSerGluPheProAspLysPheSerCysValProHisThrThrArgProLysArg 571
 Qy 1390 AAATATGAGNAGGCTTTGACCCCTACTCTATGTTCCACCCAGCAGACATCATCGGAAG 1449
 Db 572 GluTyrGluIleAspGlyArgAspTyrHisPheValSerSerArgGluLysMetGluLys 591
 Qy 1450 GATCTCGGCTCCTACCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAGAGCGGT 1509
 Db 592 AspIleGlnAlaHisLysPheIleGluAlaGlyGlnTyrAsnSerHisLeuTyrGlyThr 611
 Qy 1510 GTGACTCCCCCAT-----GGGAAG-----GTGCTC 1536
 Db 612 SerValGlnSerValArgGluValAlaGluGlnGlyLysHisCysIleLeuAspValSer 631

Qy 1537 GTTCTCTGTGTATGAGCGGGAGCTGCTGAGCGCATGCTGGCATTTGT----- 1587
 Db 632 AlaAsnAlaValArgArgLeuGlnAlaAlaHisLeuHisProIleAlaIlePheIleArg 651
 Qy 1588 -----AAAGGGACGAGATCATGCGCAATCAACGGCAAGATGTGACAGACTACACCTG 1641
 Db 652 ProArgSerLeuGluAsnValLeuGluIleAsnLysArgIleThrGluGluGlnAlaArg 671
 Qy 1642 GCTGAGCTGACCTGCC 1659
 Db 672 LysAlaPheAspArgAla 677

RESULT 8

Z03_CANFA STANDARD; PRT; 898 AA.
 AC 062683;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
 DE occludens 3 protein) (Tight junction protein 3).
 GN TP3 OR ZO3
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=98198478; PubMed=9531559;
 RA Haskins J., Gu L., Wittchen E.S., Hibbard J., Stevenson B.R.;
 RT "ZO-3, a novel member of the MAGUK protein family found at the tight
 RT junction, interacts with ZO-1 and occludin";
 RL J. Cell Biol. 141:199-208(1998).
 CC -1- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL: AF023617; AAC39177.1; .
 DR HSSP: P31016; 1BFE.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00825; Guanylate_kin; 1.
 DR SMART: SM00072; GUKC; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 DR KW Tight junction; SH3 domain; Repeat; Membrane.
 FT DOMAIN 11 93
 FT DOMAIN 187 264 PDZ 1.
 FT DOMAIN 369 435 PDZ 2.
 FT DOMAIN 464 538 PDZ 3.
 FT DOMAIN 650 750 GUANYLATE KINASE.
 SQ SEQUENCE 898 AA; 98414 MW; 8091D6132DB9F15D CRC64;

Alignment Scores:

Pred. No.: 2,07e-06 Length: 898
 Score: 249.50 Matches: 149

Percent Similarity: 37.12% Conservative: 83
 Best Local Similarity: 23.84% Mismatches: 207
 Query Match: 6.06% Indels: 186
 DB: 1 Gaps: 31

US-09-502-945-4 (1-2236) x 203_CANFA (1-998)

```

Qy 385 GGCCTCGCCTGAGTGGTGGTGGCTGGCTGGG---CTCTTCATCTCC 441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 GlyPheGlyIleAlaIleSerGlyArgAspArgAlaSerGlySerValValSer 40
Qy 442 CACCTCATCAAGCGGTCAGCACACAGCTCGGCTCCAGGTAGGAGGAGATCTGC 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 AspValValProGlyGlyProAlaAspGly---ArgLeuGlnThrGlyAspHisVal 59
Qy 502 CGGATCAATGATATCTCATCTCCCTGTACCCATGAGGAGTTCATCAACTTCGA 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 MetValAsnGlyValSerMetGluSerValThrSerPheAlaIleGlnIleLeuLys 79
Qy 562 ACC---AAGAAACTGTCTCATCAAAAGTG-----AGACATCGGCTGATCCCC 609
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 ThrCysThrLysLeuAlaAsnIleThrValLysArgProArgLysIleGlnLeuProAla 99
Qy 610 CTGAAGACTCT------CCTGATGAGCCCTCATCTGG 642
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 ThrLysAlaGlyThrSerGlyArgGlyGlnGlyLeuGluGluGluAlaAspCysGly 119
Qy 643 CAG---TATGTGCATCAGTTTGTGCGAATCTGGG----- 675
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 GlnGlyTyrAspGlyAspThrSerSerGlySerGlyArgSerTrpAspLysArgSerArg 139
Qy 676 -----GGCGTCGAGGCGAGCTGGCTCCCTCGGAAATCGGAA- 714
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140 ArgAlaArgThrGlyArgArgAsnGlnAlaGlySerArgGlyArgArgSerProGlyGly 159
Qy 715 ACACAGGAG----- 723
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 AsnSerGluAlaAsnGlyLeuAlaLeuValSerGlyPheLysArgLeuProArgGlnAsp 179
Qy 724 -----AAGAGGTCTTCATCAGCTCGGTAGGCTCCGAGGCGCTGGC 765
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 ValHisMetArgProValLysSerValLeuValArgArgThrGluSerGluGluPheGly 199
Qy 766 TCAGCATTTTCAGCGGCGCCCATCCAGAGCCTGTCATCTTTATCAGCCATGTGAACCT 825
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 ValThrLeuGlySer-----GlnIlePheIleLysHisIleThrAsp 213
Qy 826 GCCTCCCTCTCTGCT---GAGTGGGATTGGAGATAGGGACCATGTCGAAGTCAAT 882
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 SerGlyLeuAlaAlaArgAsnArgGlyLeuGlnGlyAspLeuIleLeuGlnIleAsn 233
Qy 883 GCGCTCGACTTCTTAACCTGGATCACAGAGGAGGCTGTAATGTGCTGAAATAATAGCCGC 942
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 GlyValSerSerGluAsnLeuSerLeuSerAspThrArgArgLeuIleGluLysSerGlu 253
Qy 943 ----AGCCTGACCATCTCCATTAGTCGAGCTGGCGCGGAGCTGTTTCATCAGACCGG 999
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 GlyLysLeuThrLeuLeuValLeuArgAspArgGlyGlnPheLeuVal--AsnIleProp 273
Qy 1000 GAGCGGCTGGCAGCGCGGCGAGCTGAGCT-----GCAGCGGAGGAG 1044
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 roAlaValSerAspSerAspSerAspSerPheLeuAspIleSerAlaLeuGlyS 293
Qy 1045 CTCTCATCGAAGCGGCTGGCGATGGAGTCCCAACAGATCTCCAGGAGCAGGAG 1104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 erGluLeuSerGlnAlaValProSerHisValProProProProHisAla----- 310
Qy 1105 ATGGAGCGGCAAGGAGAAAGAAATTCGCCAGAGGCGAGCAGAGAAATGAGATAC 1164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 -----GlnArgSerLeu----- 314
Qy 1165 CGGAGGAGATCGACAGATTCT---AGAGGAGGAGAGAGAGTTTAAGAGCAATGGGAA 1221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 315 --AspSerAspGlyThrAspSerProArgAspSerProProLeu-----A 329
Qy 1222 GAAGACTGGGCTCAAAAGAACAGCTACTCTTCCTAAAACCATCACTGCTGAGGTAC 1281
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 rArgGluAsnSerLeuAspSerArgThrIleSer-----GluP 342
Qy 1282 CCAGTACCCCTTCGCAAGCCAAAGTATGATCAGGAGTGGAGCTGAGCTCGAGCCGCA 1341
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 342 roAspAlaPro-ArgHisSerSerTyrArgValProSerSerGlnSerAla 361
Qy 1342 GATCAGCTGGATGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 GluAsp-----Arg 364
Qy 1402 GGCTTTGACCCCTACTCTATGTTTACCCACAGCAGATCATGGGAAGAGTATGCTCGGCTC 1461
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 GlyTyrSerPro-----AspSerArgVal 372
Qy 1462 CTAGCATCAAGAGGAGGAGGATCCTTAGACCTGGCCCTGGAAGCGGTGTGGACTCCCC 1521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 ValArgPheHisLysGlyThrThrIleGlyLeuArgLeuAlaGlyGlyAsnAsp----- 390
Qy 1522 ATTGGGAAGGTGGTCTGCTGTCTGTATGAGCGGGAGCTGTGAGCGGCATGTGTGGC 1581
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 ValGly---IlePheValSerGlyValGlnGlySerProAlaAspGlyGln---Gly 408
Qy 1582 ATTGTGAAGGGGACGAGATCATGCAATCAACGGCAAGATTTGTACAGACTACACCTG 1641
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 IleGlnGluGlyAspGlnIleLeuGlnValAsnAspValProPheArgAsnLeuThrArg 428
Qy 1642 GCTGAGCTGACGTGCCCTGTCAGAACGCTGGAATCAGGGCGGGGACTGGATCGACTT 1701
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 GluGluAla-----ValGlnPhe 434
Qy 1702 GTGCTTCGCTGCTCCGCCCAAGAGAGTATGACGATGAGCTGACCTTCTTCTCAAGTCC 1761
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 LeuValAlaLeuProProGlyGluGluValGlu-----LeuValThr 448
Qy 1762 AAAAGGGGAAACCAAAATTCACGCTTAGGAAACAGTACGCTCCGCGCCCTCGTCAAC 1821
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 449 GlnArg-----AsnGluAspIlePheArgLysMetValGln 460
Qy 1822 ACAAGCCTCGGACCGCTTGTAGAGAGGCGCATACACACACACAGATGCGATCCTTGG 1881
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 SerArgVal-----GlyAspSerPheTyrIleArgThrHis---Phe 473
Qy 1882 GACCTGAATCTATCACCCAGGAATCTCAAACTCCTTGGCCCTGAACAGGCGGACAGATA 1941
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 GluLeuGluAlaSerPro-----ProSerGlyLeu-----GlyPheThr 486
Qy 1942 AGGAACAGCTCGGCGCACTTTTGAAGCCCAATGTGGAGGAAAGGGA----- 1989
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 ArgGlyAspValPheHisValLeuAspThrLeuCysProGlyProGlyProSerGlyAla 506
Qy 1990 -----GCAGCCAGCGGTTGGAGAGAAATCTCAAGGATCCAGACTCTCAT 2034
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 ArgGlyThrHisTrpLeuAlaValArgMetGlyArgAspLeuArgGlnGluArgGly 526
Qy 2035 TCCTTCTCT 2043
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 527 IleIlePro 529
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 9

DLG1_HUMAN

ID DLG1_HUMAN STANDARD; PRT; 904 AA.

AC Q12959; Q12958;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Presynaptic protein SAP97 (Synapse-associated protein 97) (Discs,

DE large homolog 1).

GN DLG1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95024052; PubMed=7937897;
 RA Lue R.A., Marfatia S.M., Branton D., Chishti A.H.;
 RT "Cloning and characterization of hdlg: the human homologue of the
 RT Drosophila discs large tumor suppressor binds to protein 4.1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).
 RN [2]
 RX X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 460-555.
 RA MEDLINE=96338231; PubMed=8757139;
 RA Cabral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F.,
 RA Marfatia S.M., Chishti A.H., Liddington R.C.;
 RT "Crystal structure of a PDZ domain.";
 RL Nature 382:649-652(1996).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. ASSOCIATES WITH PROTEIN 4.1.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U13897; AAA50599.1; -;
 DR EMBL; U13896; AAA50598.1; -;
 DR PDB; 1PDR; 23-JUL-97.
 DR Genew; HGNC:2900; DLG1.
 DR MIM; 601014; -;
 DR InterPro; IPR000619; Guanylate_kin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00018; SH3; 1.
 DR Pfam; PF00595; PDZ; 3.
 DR Pfam; PF00625; Guanylate_kin; 1.
 DR Prodom; PD000066; SH3; 1.
 DR SMART; SM00072; GUKC; 1.
 DR SMART; SM00228; PDZ; 3.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS50106; PDZ; 3.
 DR PROSITE; PS50002; SH3; 1.
 DR SH3 domain; Repeat; Alternative splicing; 3D-structure.
 DR DOMAIN 224 310 PDZ 1.
 DR DOMAIN 319 405 PDZ 2.
 DR DOMAIN 466 546 PDZ 3.
 DR DOMAIN 581 651 SH3.
 DR DOMAIN 714 904 GUANYLATE KINASE.
 DR VARSPLIC 669 680 EIPDMSGKGLK -> QSFNDKRNKLNFSRKFPPYKNDQS
 FT EQETSDAO (IN ISOFORM 2).
 FT
 SQ SEQUENCE 904 AA; 100354 MW; B78798D6BB0920D4 CRC64;

Alignment Scores:

Pred. No.:	3,86e-06	Length:	904
Score:	244.50	Matches:	121
Percent Similarity:	38.88%	Conservative:	87
Best Local Similarity:	22.62%	Mismatches:	202
Query Match:	5.93%	Indels:	125
DB:	1	Gaps:	22

US-09-502-945-4 (1-2236) x DLG1_HUMAN (1-904)

QY	340	TCAGGAGCTGAAGGAGGTGGCTGCGACCGCTGCGACCCCGGAGCCCTCGCCTGAGT	399
DB	314	SerGluLysIleMetGluIleLysLeuIleLys---GlyProLysGlyLeuGlyPheSer	332
QY	400	GTGGTGGTGGC-----CTGGAGTTTGGCTGGCTTCATCCTCCAC	444
DB	333	IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys	352
QY	445	CTCATCAAGCGCGTCAGCAGCAGCGTCGGG---CTCCAGGTAGGGAGCAGATCGTC	501
DB	353	IleIleGluGlyGlyAlaAlaHisLysAspGlyLysLeuGlnIleGlyAspLysLeuLeu	372
QY	502	CGGATCAATGGATATCCATCTCTCTCTGATACCATCAGGAGGTCATCACTTCGGA	561
DB	373	AlaValAsnAsnValCysLeuGluGluValThrHisGluGluAlaValThrAlaLeuLys	392
QY	562	---ACCAAGAAAGACTGTGTCATCAAAAGTG-----AGACACATC	597
DB	393	AsnThrSerAspPheValTyrLeuLysValAlaLysProThrSerMetTyrMetAsnAsp	412
QY	598	GGCTGATGCC-----GTGAAAGCTCTCTGATGAGCCCTCACTTGGCAGTATGTG	651
DB	413	GlyTyrAlaProProAspIleThrAsnSerSerSerGlnPro-----Val	427
QY	652	GATCAGTTTGTGCGGAATCTGGGGCGTCGGAGGAGCCTGGCTCCCTC-----	702
DB	428	AspAsnHisValSerProSerSerPheLeuGlyGlnThrProAlaSerProAlaAspTyr	447
QY	703	-----GGAAATCGGGAA---AACAGGAGGAGGAAGGTCTTC	735
DB	448	SerProValSerLysAlaValLeuGlyAspGluIleThrArgGluProArgLysVal	467
QY	736	ATCAGCCTGTGAGTCCCGAGCGCTTGGTCAGCAATTTCCAGCGGCCCTCCAGAAG	795
DB	468	ValLeuHisArgGlySerThrGlyLeuGlyPheAsnIleValGlyGlu---AspGly	486
QY	796	CCTGGCATCTTATCAGCCATGTG-----AACTGGCTCCCTGTCTGTGAGGTG	846
DB	487	GluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu---	505
QY	847	GGATTGGAGATAGGGGACCACTGTGCAAGTTCAGTGGCTGCAGCTTCTCTAACCTGGAT	906
DB	506	---LeuArgLysGlyAspArgIleIleSerValAsnSerValAspLeuArgAlaAlaSer	524
QY	907	CACAAGAGGCTGTAATGTGCTGAAATAGCCGAGCCTGACCATCTCCATTTAGCT	966
DB	525	HisGluGlnAlaAlaAlaLeuLysAsnAla---GlyGlnAlaValThrIleValAla	543
QY	967	GCAGCTGGCGGAGCTGTTTCATGACAGACCGGCGCTGGCAGAGCGCGGCGAG---	1023
DB	544	GlnTyrArgProGluGluTyrSerArgPheGluAlaLysIleHisAspLeuArgGluGln	563
QY	1024	-----CGTGAGCTGCGAGCGCGAGGAG	1044
DB	564	MetMetAsnSerSerIleSerSerGlySerLeuArgThrSerGlnLysArgSer	583
QY	1045	CTTCTCATGACAGAGCGGCTGGCGATGAGTGCACCAACAAAG-----	1083
DB	584	LeuTyrValArgAlaLeuPheAspTyrAspLysThrLysAspSerGlyLeuProSerGln	603
QY	1084	-----ATCCTC-----	1089
DB	604	GlyLeuAsnPheLysPheGlyAspIleLeuHisValIleAsnAlaSerAspGluTrp	623
QY	1090	---CAGGAGCAGCAGGAGATGGAGCGGCAAGGAGAAAGAAATT-----GCC	1134
DB	624	TrpGlnAlaArgGlnValThrProAspGlyGluSerAspGluValIleProSer	643
QY	1135	CAGAAGGCAGCAGAGGAAATGAGATACCGGAAGGAGATGGAACAGATTGTAGAGGAG	1194
DB	644	LysArgArgValGluLysLysGluArgAlaArgLeuLysThrValLysPhe---AsnSer	662
QY	1195	GAAGAGAAGTTTAAGAAACAATGGGAAGAGACTGGGCTCAAAG-----	1239


```

Db 663 LysThrArgAspLysGlyGluLeuProAspMetGlySerLysGlyLeuLysHisVal 682
Qy 1240 -----
Db 683 ThrSerAsnAlaSerAspSerCluSerSerTyrArgGlyGlnGluGluTyrValLeuSer 702
Qy 1255 CCTAAACCATCAGCTGCTGAG-----GTCACCCAGTACCCCTTCGCAAGCCA 1302
Db 703 TyrGluProValAsnGlnGlnValAsnTyrThrArgProValIleIleLeuGlyPro 722
Qy 1303 AAGTATGATCAGGAGTGAACCTGAGCTCGAGCCCGAGATGAGCTGGATGGAGGCACG 1362
Db 723 MetLysAspArgIleAsnAspLeuIleSerGluPheProAspLysPheGlySerCys 742
Qy 1363 GAGGACAGGAGCAGCATTTCCGAAATATGAGGAAGGCTTTCAGCCCTACTCTCTATG 1422
Db 743 ValProHisThrThrArgProLysArgAspTyrGluValAspGlyArgAspTyrHisPhe 762
Qy 1423 TTCACCCAGCAGCATGATGGGAGGAGTTCGGCTCCCTACGATCAAGAAGAGGGA 1482
Db 763 ValThrSerArgGluGlnMetGluLysAspIleGlnGluHisLysPheIleGluAlaGly 782
Qy 1483 TCCTTAGACTGGCCCTGGAAGCGGTGTGGACTCCCTCCCATTTGGGAAGGTGGCTTTCT 1542
Db 783 GlnTyrAsnAsnHisLeuTyrGly-----ThrSerValGln 794
Qy 1543 GCTGTGTATGACGGGGAGCTGCTGAGCGGCATGTGGCATTTGTAAGAGGACGAGATC 1602
Db 795 SerValArgGluValAlaGlyLysGlyLysHisCys-----Ile 807
Qy 1603 ATGGCAATCAACGGCAAGATTGTGACAGACTACACCCGCTGAG 1647
Db 808 LeuAspValSerGlyAsnAlaIleLysArgLeuGlnIleAlaGln 822

RESULT 10
DLG3_MOUSE
ID DLG3_MOUSE STANDARD; PRT; 849 AA.
AC P01175;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic protein SAPI02 (Synapse-associated protein 102) (Discs,
DE large homolog 3).
GN DLG3 OR DLG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Kohmura N., Makino S., Yagi T.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC SUBUNIT NP2B.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D87117; BAA13249.1; .
DR HSSP; Q12959; 1PDR.
DR MGD; MGI:1888986; Dlg3.
DR InterPro; IPR000619; Guanylate_kin.

```

```

DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00072; GukC; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS0106; PDZ; 3.
DR PROSITE; PS50002; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50002; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 149 235 PDZ 1.
FT DOMAIN 244 330 PDZ 2.
FT DOMAIN 404 484 PDZ 3.
FT DOMAIN 519 589 SH3.
FT DOMAIN 659 849 GUANYLATE_KINASE.
SQ SEQUENCE 849 AA; 93482 MW; EF3EF2D7513538EE CRC64;

Alignment Scores:
Pred. No.: 4,6e-06 Length: 849
Score: 243.00 Matches: 137
Percent Similarity: 36.29% Conservative: 96
Best Local Similarity: 21.34% Mismatches: 191
Query Match: 5.90% Indels: 218
DB: 1 Gaps: 29

US-09-502-945-4 (1-2236) x DLG3_MOUSE (1-849)
Qy 8 CGGTGCGGTGCGGCTCTTT-----CCAGTCTCTGGCAGCCGG 46
Db 70 ArgThrLysAlaLysLeuIleProThrGlyArgAspValGlyProValProLysPro 89
Qy 47 GCACCCGAGCAACGGTCTGTCGCAACGACGACGCTGACCTGGCCGCCA---TGGACC 103
Db 90 ValPro-----GlyLysSerThrProLysLeuAsnGlySerGlyProGlyTyr--- 105
Qy 104 GAAAGTGGCCGAGAAATTCGGCATAAGGTGGATTTTCTGATTGAAATGATGACAGAGA 163
Db 106 -----TrpProGlu----- 108
Qy 164 AGGACTATCTCTATGATGTCGTGCGAATGTACACACGACCATGGACGTGGCCGTGCTG 223
Db 109 -----CysThrAsnArgAspTyr----- 116
Qy 224 TGGGAGACCTGAAGCTGGTGCATCAATGAACCCAGCCGCTCTGCTCTGTTGATGCCATTC 283
Db 117 -----TyrGluIleAlaSerProAlaPro---LeuLeuValAsnPro-Gl 130
Qy 284 GCGCGCTGATCCCACTGAAGCACCAGGTGGAAATATGATCAGCTGACCCCGCGCTCCA 343
Db 130 uAlaLeuGluProSerLeuSerValAsnGly-SerAspGlyMetPhe----- 145
Qy 344 GGAAGCTGAAGAGGTGGTCTGACCGCTGACCCCGAAGCGCTGGCTGAGTGTGC 403
Db 146 --LysTyrGluGluIleValLeuGluArgGlyAsn---SerGlyLeuGlyPheSerIleA 164
Qy 404 GTGTGGCTCGAG-----TTTGGCTGTGGCTCTTCATCTCCACCTCA 448
Db 164 laGlyIleAspAsnProHisValProAspProGlyIlePheIleThrLysIleI 184
Qy 449 TCAAGGCGGTGACGACAGACGCTCGGG---CTCCAGGTAGGGGAGAGATCTCGGA 505
Db 184 leProGlyGlyAlaAlaAlaMetAspGlyArgLeuGlyValAsnAspCysValLeuArgV 204
Qy 506 TCAATGATATTCATCTCTCTGTCCTGTCACCCATGAGGAGGTGATCAACCTCATTCGACCA 565
Db 204 alAsnGluValAspValSerGluValHisSerArgAlaValGluAlaLeuLysGluA 224
Qy 566 AGAAACTGTGTCTCCATCAAGTGAGACACATCGCGCTGATCCCGCTGAAAAGCTCTCTCTG 625

```

```

Db 224 laGlyProVal-----ValArg-----LeuValValArgArgGlnP 237
Qy 626 ATGAGCCCTCACTGGCAGTATGTCATGTTGTGTGCGAATCTGGGCGCTGCAG 685
Db 237 roProGluThrIleMetGluValAsn----- 246
Qy 686 GCAGCGCTGGCTCCCTCGAAATCGGAAACAAGGAGAAGAGCTTTCATCAGCCTGG 745
Db 247 -----LeuLeuL 249
Qy 746 TAGGTCCTCGAGGCTTGCTGCATGCTTCCAGCGGCCCTCATCAGAG-----CTG 799
Db 249 ysGlyProLysGlyLeuGlyPheSerIleAlaGlyIleGlyAsnGlnHisIleProG 269
Qy 800 GC-----ATCTTTATCACCCTGTCAACCTGGCTCCCTGCTGCTGAGGTGGA- 849
Db 269 lyAspAsnSerIleThrIleThrLysIleIleGlyGlyAlaAlaGlnLysAspGlyA 289
Qy 850 --TTGAGATAGGACGATGTTCGAAGTCAATGGGCTGCATCTCTTAACCTGGATC 907
Db 289 rgLeuGlnIleGlyAspArgLeuLeuAlaValAlaAsnThrAsnLeuGlnAspValArgH 309
Qy 908 ACAAGAGGCTTAATGTGCTGAATAATAGCCGAGCTGACCATCTCCATGTAGCTG 967
Db 309 IsGluGluAlaValAlaSerLeuLysAsnThrSerAspMetValTyrLeuLysValAlaL 329
Qy 968 CAGCTGGCGGGAGCTGTTTCATGACAGACCGGGAGCGGTGTCAGAGCGCGGACGCTG 1027
Db 329 ysProGly-----S 332
Qy 1028 AGCTGCAGCGGAGGAGCTTCTCATGCAGAGCGGTGCGGATGGAGTCCAACAAGATCC 1087
Db 332 erIleHisLeuAsnAspMetTyrAlaProAspTyrAlaSerThrPheThrAlaLeu- 351
Qy 1088 TCCAGGACGAGGATGGAGCGGCAAGAGAGAAATGCCAGAGAGGCGAGCAG 1147
Db 352 -----AlaA 353
Qy 1148 AGGAAATGAGAGATACCGGAGGAGATGGACATGTTAGAGGAGGAGAGAGATTGA 1207
Db 353 spAsnHisIleSerHisAsnSerSerLeuGlyTyrLeuGlyAlaValGluSerLysVal- 372
Qy 1208 AGAGCAATGGGAAGAAGACTGGGGCTCAAGGAACAGCTACTCTTCCTTAAACCATCA 1267
Db 372 ----- 372
Qy 1268 CTGCTGAGGTACACCACTACCCCTTCGCAAGCCAAAGTATGATCAGGAGTGGAACTG 1327
Db 373 -----ThrTyrProAlaPro-----ProG 379
Qy 1328 AGCTGAGCGCGCAGATGACCTGGATGGAGGACGCGGAGCAGGAGCAGGAGTTTC 1387
Db 379 InValProProthr----- 383
Qy 1388 GGAATATGAGGAGGCTTTGACCCCTACTCTATGTTCCACCCAGCAGCAGATCATGGGA 1447
Db 384 --ArgTyrSerPro-----IleProArgHisMetLeuAlaGluGluAspPheThrArg 401
Qy 1448 AGGATGTCGGCTCTCAGCATCAAGAGAGGAGGATCCTTAGACCTGCGCCTGGAAGCG 1507
Db 401 luProArgLysIleIleLeuHisLysGlySerThrGlyLeuGlyPheAsnIleValGlyG 421
Qy 1508 GTGTGAGTCCCTCCATTTGGGAGGTGGTCTTCTGCTGTGTATGAGCGGGAGCTGCTG 1567
Db 421 lyGluAspGlyGlu-----GlyIlePheValSerPheIleLeuAlaGlyGlyProAlaA 439
Qy 1568 AGCGCATGGGGCATTTCTGAAGAGGAGCAGATCATGGCAATCAACCGCAGATTTGTA 1627
Db 439 spLeuSerGlyGluLeuArgArgGlyAspArgIleLeuSerValAsnGlyValAsnLeuA 459
Qy 1628 CAGACTACACCTGGCTGAGGCTGACGCTGCGCTCAGAGGCGCTGGAATCAGGCGCGGG 1687
Db 459 rgAsnAlaThrHisGluGlnAlaAlaAlaLeuLysArgAla-----GlyG 475

```

```

Qy 1688 ACTGATCGACCTTGTTGTTCCCTGCGCCCAAGAGGATATGACGATGAGTGACCT 1747
Db 475 InSerVal---ThrIleValAlaGlnTyrArgProGluGluTyr----- 488
Qy 1748 TCTTCTGAAGTCCAAAGAGGGAACCAAAATTCACGCGTTAGGAAACAGTACGCTCCGC 1807
Db 489 -----SerArgPheGluSerLysIleHis-----AspLeuArgG 500
Qy 1808 CCCACCTCGTGAACACAAAGCCT-----CGGACCAAGCTTGAGA 1846
Db 500 luGlnMetMetAsnSerSerMetSerSerGlySerGlySerLeuArgThrSerGluLysA 520
Qy 1847 GA 1848
Db 520 rg 520

RESULT 11
ID_DLG3_RAT      STANDARD;      PRT;      849 AA.
AC Q62936; P70547;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic protein SAP102 (Synapse-associated protein 102) (PSD-
DE 95/SAP90 related protein 1) (Discs, large homolog 3).
GN DLG3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE=Brain;
RX MEDLINE=96374358; PubMed=8780649;
RA Mueller B.M., Kistner U., Kindler S., Chung W.J., Kuhlendahl S.,
RA Fenster S.D., Lau L.-F., Veh R.W., Haganir R.L., Gundelfinger E.D.,
RA Garner C.C.;
RT "SAP102, a novel postsynaptic protein that interacts with NMDA
RT receptor complexes in vivo.";
RL Neuron 17:255-265(1996).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RA Irie M., Hata Y., Takai Y.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC SUBUNIT NR2B.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U50147; AAA93031.1; -.
DR EMBL; U53367; AAB48561.1; -.
DR HSP; Q12959; IPDR.
DR InterPro; IPR000619; Guanylate_Kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00072; GUKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.

```



```

Db 585 ValArgAlaLeuPheAspTyrAspLysThrLysAspSerGlyLeuProSerGlnGlyLeu 604
QY 1083 -----
Db 605 AsnPhelYspGlyAspIleLeuHisValIleAsnAlaSerAspAspGluTrpTrpGln 624
QY 1084 -----
Db 625 AlaArgGlnValThrProAspGlyGluSerAspGluValGlyValIleProSerLysArg 644
QY 1102 GAGATGGAGCGGCAAGGAGAAAGAAATGTCACCAAGGACGACGAGGAATATGAGACA 1161
Db 645 ArgValGluLysLysGluArgAlaArgLeu-----LysThrValLysPheAsnSerLys 662
QY 1162 TACCGAAGGAGATGGAACAGATTGTAGAGGAG----- 1194
Db 663 ThrArgLysLysGlyGluIleProAspAspMetGlySerLysGlyLeuLysHisVal 682
QY 1195 -----
Db 683 ThrSerAsnAlaSerAspSerGluSerSerTyrHisGluTyrGlyCysSerLysGlyGly 702
QY 1234 TCAAGGAACAGTACTCTTCCTTAACCATCACTCCTCAG-----GTACAC 1281
Db 703 GlnGluGluTyrValLeuSerTyrGluProValAsnGlnGlnGluValAsnTyrThrArg 722
QY 1282 CCAGTACCCCTCGCAAGCCAAAGATATGATCAGGAGTGGAACTGAGCTCGAGCCCGCA 1341
Db 723 ProValIleIleLeuGlyProMetLysAspArgValAsnAspAspLeuIleSerGluPhe 742
QY 1342 GATGACCTGATGAGGACGAGGAGGAGGAGGAGGAGGATTCGCGGAATATGAGGAA 1401
Db 743 ProAspLysPheGlySerCysValProHisThrThrArgProLysArgAspTyrGluVal 762
QY 1402 GGCTTTGACCCCTACTCTATGTTTACCCAGACGACGATCATGGGAGGATGCCGCTC 1461
Db 763 AspGlyArgAspTyrHisPheValThrSerArgGluGlnMetGluLysAspIleGlnGlu 782
QY 1462 CTAGCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAAGGC---GGTGTGGACTCC 1518
Db 783 HisLysPheIleGluAlaGlyGlnTyrAsnAsnHisLeuTyrGlyThrSerValGlnSer 802
QY 1519 CCCATGGGAGGTGGTGGTTCCTGCTGTGTATGAGCGGGGAGCTGCTGACGGCGATGGT 1578
Db 803 -----ValArgAlaValAlaGluLysGly-----LysHisCys 813
QY 1579 GGCATTGTGAAGGGGAGGATCATGCGCAATCAACGCAAGATGTGACAGACTACACC 1638
Db 814 -----IleLeuAspValSerGlyAsnAlaIleLysArgLeuGln 826
QY 1639 CTGGCTGAG 1647
Db 827 IleAlaGln 829

RESULT 13
DLG3_HUMAN STANDARD; PRT; 817 AA.
AC Q92796; O9ULI8;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic protein SAP102 (Synapse-associated protein 102).
DE (Neuroendocrine-DLG) (NE-DLG) (Discs, large homolog 3).
GN DLG3 OR KIAA1232.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=97332623; PubMed=9188857;
RA Makino K., Kuwahara H., Masuko N., Nishiyama Y., Morisaki T.,

```

```

RA Sasaki J., Nakao M., Kuwano A., Nakata M., Ushio Y., Saya H.;
RT "Cloning and characterization of NE-dlg: a novel human homolog of the
RT Drosophila discs large (dlg) tumor suppressor protein interacts with
RT the APC protein.";
RL Oncogene 14:2425-2433(1997).
RN [2]
RP SEQUENCE OF 330-817 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
CC 1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC SUBUNIT NR2B (BY SIMILARITY).
CC 1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC 1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC 1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC 1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U49089; AAB61453.1; -.
CC EMBL; AB033058; BAA86546.1; -.
CC HSSP; Q12959; 1PDR.
CC Genew; HGNC:2902; DLG3.
CC MIM; 300189; -.
CC InterPro; IPR000619; Guanylate_kin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00595; PDZ; 3.
CC Pfam; PF00625; Guanylate_kin; 1.
CC ProDom; PD000066; SH3; 1.
CC SMART; SM00072; GUKC; 1.
CC SMART; SM00228; PDZ; 3.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
CC PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
CC PROSITE; PS50106; PDZ; 3.
CC PROSITE; PS50002; SH3; 1.
CC SH3 domain; Repeat.
KW FT DOMAIN 130 217 PDZ 1.
FT DOMAIN 226 311 PDZ 2.
FT DOMAIN 379 465 PDZ 3.
FT DOMAIN 503 568 SH3.
FT DOMAIN 628 803 GUANYLATE KINASE.
FT CONFLICT 330 381 FTALADNHSHNSLSGLYGAVESKYSVPAPQVPPTKVSPI
FT PRHMLAEEDFT -> AARRRGAMERARKFSGSLAMGLGS
FT ASASARRASORWAMPLSLRPGGDA (IN REF. 2).
FT DFGGLSDDYGAKNL -> SIKTKRKKSFLSLRKPFYKSK
FT ENMAQESSIQGGYTSNTSDESSS (IN REF. 2).
SQ SEQUENCE 817 AA; 90344 MW; 3D7512EC4713FC4E CRC64;

Alignment Scores:
Pred. No.: 1,32e-05 Length: 817
Score: 234.50 Matches: 164
Percent Similarity: 37.37% Conservative: 123
Best Local Similarity: 21.35% Mismatches: 276
Query Match: 5.69% Indels: 206
DB: 1 Gaps: 35

US-09-502-945-4 (1-2236) x DLG3_HUMAN (1-817)
QY 8 CGGTGCGGTGCGGGCTCTTTCCAGCTCTTCCAGCTGCGAGCGGGCACCC-----GAA 55

```

```

Db 70 ArgThrLysAlaLysLeuIleProThrGlyArgAspValGlyProValProLeuLysPro 89
|||.....||| |||
Qy 56 GGAAGGGTCGTGCAACGACGACGTGGACCTGGCCAGCCA---TGGACCGAAAGTGG 112
|||.....||| |||
Db 90 ValProGlyLysSerThrProLysLeuAsnGlySerGlyProSerTrp-----Trp 106
|||.....||| |||
Qy 113 CCCGAG----- 118
|||
Db 107 ProGluCysThrCysThrAsnArgAspTrpTyrGluInValAsnGlySerAspGlyMet 126
|||
Qy 119 -----AATCCCGCATAGCTGCATTTCTCG- 144
|||.....||| |||
Db 127 PheLysTyrGluIleValLeuGluArgLysAsnSerGlyLeuGlyPheSerIleAla 146
|||.....||| |||
Qy 145 -----ATTGAAATGATGCAGAGAGGAC-----TATCTCTATGATGCTGCTG 186
|||.....||| |||
Db 147 GlyGlyIleAspAsnProHisValProAspAspProGlyIlePheIleThrLysIleIle 166
|||.....||| |||
Qy 187 CGAATGTACCAACGACCATGGACGTGGCGTCTGCTGGGAGACCTGAAGCTGTATC 246
|||.....||| |||
Db 167 ProGlyGlyAlaAlaAlaMetAspGlyArgLeuGlyValAsnAspCysValLeuArgVal 186
|||.....||| |||
Qy 247 AATGAACCCAGCGCTCTGCTCTGTTGATGCCATT-----CGCCGCTGATCCCACTG 300
|||.....||| |||
Db 187 AsnGlu-----ValGluValSerGluValValHisSerArgAlaValGluAlaLeu 203
|||.....||| |||
Qy 301 AAGCACCAAGGTGAATATGATCAGTGCACCCCGCGCTCCAGG-----AAG 348
|||.....||| |||
Db 204 LysGluAlaGlyProValValArgLeuValValArgArgGlnProProGluThr 223
|||.....||| |||
Qy 349 CTGAAGGAGGTGGCTCTGGACCGCTGCACCCGAGCCCTGGCTGAGTGTGGTGGT 408
|||.....||| |||
Db 224 IleMetGluValAsnLeuLys---GlyProLysGlyLeuGlyPheSerIleAlaGly 242
|||.....||| |||
Qy 409 GGC-----CTGAGTTTGGCTGTGGCTCTTCANCTCCACCTCATCAAA 453
|||.....||| |||
Db 243 GlyIleGlyAsnGlnHisIleProGlyAspAsnSerIleIleThrLysIleIleGlu 262
|||.....||| |||
Qy 454 GCGGTTCAGGACGACAGCGTCGGG---CTCCAGGTAGGGACGAGATGCTCCGGATCAAT 510
|||.....||| |||
Db 263 GlyGlyAlaAlaGlnLysAspGlyArgLeuGlnIleGlyAspArgLeuAlaValAsn 282
|||.....||| |||
Qy 511 GGATATTCATCTCTCTGTACCATGAGGAG---GTCATCAACCTCATTCGAAACCAAG 567
|||.....||| |||
Db 283 AsnThrAsnLeuGlnAspValArgHisGluAlaValAlaSerLeuLysAsnThrSer 302
|||.....||| |||
Qy 568 AAAACTGTGTCATCAAGTGAACATCGGC-----CTGATCCCGCTGAAAAGC 618
|||.....||| |||
Db 303 AspMetValTyrLeuLysValAlaLysProGlySerLeuHisLeuAsnAspMetTyrAla 322
|||.....||| |||
Qy 619 TCTCTGTATGACCCCTCAGCTGG---CAGTATGTGGATCAGTTGTGTGCGAATCTGGG 675
|||.....||| |||
Db 323 ProProAspTyrAlaSerThrPheThrAlaLeuAlaAspAsnHisIleSerHisAsn--- 341
|||.....||| |||
Qy 676 GCGTGCAGGACGCTGGCTCCCTCGAAATCGGAAACACAG----- 720
|||.....||| |||
Db 342 -----SerSerLeuGlyTyrLeuGlyAlaValGluSerLysValSerTyrProAla 358
|||.....||| |||
Qy 720 ----- 720
|||
Db 359 ProProGlnValProProThrArgTyrSerProIleProArgHisMetLeuAlaGluGlu 378
|||
Qy 721 -----GAGAAGAAGTCTTCATCAGCTGTGTAGCTCCCGAGCGCTTGGCTGC 768
|||.....||| |||
Db 379 AspPheThrArgGluProArgLysIleIleLeuHisLysGlySerThrGlyLeuGlyPhe 398
|||.....||| |||
Qy 769 AGCATTTCCAGCGCCCATCCAGACCTGGCATCTTTATCAGCCATGTG----- 819
|||.....||| |||
Db 399 AsnIleValGlyGlyGlu---AspGlyGlyGlyIlePheValSerPheIleLeuAlaGly 417
|||.....||| |||
Qy 820 AAACCTGGCTCTCTGCTGAGTGGGATTTGGAGATAGGGACCATGTTGTCGAGTGC 879
|||.....||| |||

```

```

Db 418 GlyProAlaAspLeuSerGlyGlu-----LeuArgArgGlyAspArgIleLeuSerVal 435
|||.....||| |||
Qy 880 AATGCGTCGACTTCTCTAACTCGATCACAAGAGGCTGTAATATGCTGAAATAATAGC 939
|||.....||| |||
Db 436 AsnGlyValAsnLeuArgAsnAlaThrHisGluGlnAlaAlaAlaLeuLysArgAla 455
|||.....||| |||
Qy 940 CGCAGCCTCACCATCTCCATTGATGTGACGTGGCGGAGCTGTTTCATGACACACCGG 999
|||.....||| |||
Db 456 ---GlyGlnSerValThrIleValAlaGlnTyrArgProGluGluTyrSerArgPheGlu 474
|||.....||| |||
Qy 1000 GAGCGGCTGGCAGAGCGCGCACCGCTGACCTGCAG----- 1035
|||.....||| |||
Db 475 SerLysIleHisAspLeuArgGluGlnMetMetAsnSerSerMetSerSerGlySerGly 494
|||.....||| |||
Qy 1035 ----- 1035
|||
Db 495 SerLeuArgThrSerGluLysArgSerLeuTyrValArgAlaLeuPheAspTyrAspArg 514
|||.....||| |||
Qy 1036 ---CGGAGAGCTTCTCATGTCAGAACCGGCTGGCGATGGAGTCCAAACAAGATCCTC--- 1089
|||.....||| |||
Db 515 ThrArgAspSerCysLeuProSerGlnGlyLeuSerPheSerTyrGlyAspIleLeuHis 534
|||.....||| |||
Qy 1090 -----CAGGACGACGAGGAGTGGAGCGCAAGG 1119
|||.....||| |||
Db 535 ValIleAsnAlaSerAspGluTrpTrpGlnAlaArgLeuValThrProHisGlyGlu 554
|||.....||| |||
Qy 1120 AGAAGAAATTT-----GCCAGAAGGACGACAGAGGAAATGAGATACACCGG 1167
|||.....||| |||
Db 555 SerGluGlnIleGlyValIleProSerLysLysArgValGluLysLysGluArgAlaArg 574
|||.....||| |||
Qy 1168 AAGGAGATGAACAG-----ATTGTAGAGGAGGAGAGAGAGTTTAAG 1209
|||.....||| |||
Db 575 LeuLysThrValLysPheHisAlaArgThrGlyMetIleGluSerAsnArgAspPhePro 594
|||.....||| |||
Qy 1210 AAGCAATGGGAAGACACTGGGCTCAAG-----GAACAGCTACTC 1251
|||.....||| |||
Db 595 GlyLeuSerAspTyrTyrGlyAlaLysAsnLeuLysGlyGlnGluAspAlaIleLeu 614
|||.....||| |||
Qy 1252 TTGCTTAAACACCTCACT---GCTGAGGTACAC-----CCAGTACCCCTTCGCAAG 1299
|||.....||| |||
Db 615 SerTyrGluProValThrArgGlnGluIleHisTyrAlaArgProValIleIleLeuGly 634
|||.....||| |||
Qy 1300 CCAAGATGATCATCAGGAGTGGACCTGACCTCGACCCCGCAGATGACCTGGATGGAGGC 1359
|||.....||| |||
Db 635 ProMetLysAspArgValAsnAspAspLeuIleSerGluPheProHisLysPheGlySer 654
|||.....||| |||
Qy 1360 ACGGAGGACGAGGAGACGAGGATTTCCGAAATATGAGGAAGGCTTGTACCCCTACTCT 1419
|||.....||| |||
Db 655 CysValProHisThrThrArgProArgArgAspAsnGluValAspGlyGlnAspTyrHis 674
|||.....||| |||
Qy 1420 ATGTTTACCCAGACAGATCATGGGAGAGGATGTCCGGCTCCTACGCATCAAGAAGGAG 1479
|||.....||| |||
Db 675 PheValValSerArgGluGlnMetGluLysAspIleGlnAspAsnLysPheIleGluAla 694
|||.....||| |||
Qy 1480 GGATCTTACACCTGGCCCTGGAAGC---GGTGTGACTCCCCCATTTGGGAAGTGGTC 1536
|||.....||| |||
Db 695 GlyGlnPheAsnAspAsnLeuTyrGlyThrSerIleGlnSer----- 708
|||.....||| |||
Qy 1537 GTTCTCTGTGATGAGCGGAGCTGTGCTGAGCGCAT-----GGT 1578
|||.....||| |||
Db 709 ValArgAlaValAlaGluArgGly-----LysHisCysIleLeuAspValSerGly 725
|||.....||| |||
Qy 1579 GGCATTGTGAAAGG----- 1593
|||
Db 726 AsnAlaIleLysArgLeuGlnGlnAlaGlnLeuTyrProIleAlaIlePheIleLysPro 745
|||.....||| |||
Qy 1594 -----CAGCAGATCATGGCAATCAACGCGCAAGATTGTGACAGACTACACCTGGCT 1644
|||.....||| |||
Db 746 LysSerIleGluAlaLeuMetGluMetAsnArg-----GlnThrTyrGlu 761
|||.....||| |||
Qy 1645 GAGCTCAGCTGCCCTGCAGGAAGGCTGG-----AATCAGGCGGGGACTCGATC 1695
|||.....||| |||
Db 762 GlnAlaAsnLysIleTyrAspLysAlaMetLysLeuGlnGluPheGlyGluTyrPhe 781
|||.....||| |||

```


Db	454	ProLaAspLeuSerGlyGlu-----LeuGlnArgGlyAspGlnIleLeuSerValAsn	471
Qy	883	GGCGTCGACTTCTTAACCTGGGATCACAAGGAGGCTGTAATGCTGCTAAAAAATAGCGCG	942
Db	472	GlyIleAspLeuArgGlyAlaSerHisGluGlnAlaAlaAlaLeuLysGlyVala---	490
Qy	943	AGCCTGACCATCTCCATTGAGTCGTAGCTGCAGCTGCCGGGAGCTGTTTCATGACAGACGGGAG	1002
Db	491	GlyGlnThrValThrIleAlaGlnTyrGlnProGluAspTyrAlaArgPheGluAla	510
Qy	1003	CGCGTCGACAGGCGCGGCAG-----	1023
Db	511	LysIleHisAspLeuArgGluGlnMetMetAsnHisSerMetSerSerGlySerGlySer	530
Qy	1024	---CGTGAGCTGCAGCGGCAGGAGCTTCTCTCCAG-----	1056
Db	531	LeuArgThrAsnGlnLysArgSerLeuTyrValArgAlaMetPheAspTyrAspLysSer	550
Qy	1056	-----	1056
Db	551	LysAspSerGlyLeuProSerGlnGlyLeuSerPheLysTyrGlyAspIleLeuHisVal	570
Qy	1057	-----AAGCGCTGGCGATGGAGTGGAGTCCACACAG	1083
Db	571	IleAsnAlaSerAspAspGluTrpTrpGlnAlaArgValMetLeuGluGlyAspSer	590
Qy	1084	---ATCTCCAGGACGACGAGAGATGGAGCGCAAGGAGGAGAAAGAA--	1128
Db	591	GluGluMetGlyValIleProSerLysArgArgValGluArgLysGluArgAlaArgLeu	610
Qy	1129	-----ATGCCCAAGAGCGACGACAGAGAAAAT	1155
Db	611	LysThrValLysPheAsnAlaLysProGlyValIleAspSerLysGlySerPheAsnAsp	630
Qy	1156	GAGAGA-----TACCGGAAGGAGATGGAACAGATGTAGAGGAGGAAG	1200
Db	631	LysArgLysLysSerPheIlePheSerArgLysPheProPheTyrLysAsnLysGluGln	650
Qy	1201	AAGTTTAAGAAGCAATGGGAAGAGACTGGGCTCAAAGAACAGCTACTCTTCCTCAAA	1260
Db	651	SerGluGlnGluThrSerAspProGluArgGlyGlnGluAspLeuIleLeuSerTyrGlu	670
Qy	1261	ACCATCACTGCTCAG-----GTACACCACGATACCCCTCCGACAGCCAAAGTAT	1308
Db	671	ProValThrArgGlnGluIleAsnTyrThrArgProValIleIleLeuGlyProMetLys	690
Qy	1309	GATCAGGAGGTGGAACCTGAGCTCGAGCCGACAGATGACCTGGATGGAGCAGCGGAGG	1368
Db	691	AspArgIleAsnAspLeuIleSerGluPheProAspLysPheGlySerCysValPro	710
Qy	1369	CAGGAGAGCAGGATTCCCGAANTATGAG---GAAGGCTTTGACCCCTACTCTATGTTCT	1425
Db	711	HisThrThrArgProLysArgAspTyrGluValAspGlyArgAspTyrHisPheValIle	730
Qy	1426	ACCCAGAGCAGATCATGGGAAGAGATGCCGGCTCTTACGCATCAAGAAGGAGGATCC	1485
Db	731	SerArgGluGln---MetGluLysAspIleGlnGluHisLysPheIleGluAlaGlyGln	749
Qy	1486	TTAGACCTGGCCCTCGAAGGC---GGTGAGGACTCCCCCATTTGGAGGTGGTCTGTTCT	1542
Db	750	TyrAsnAspAsnLeuTyrGlyThrSerValGlnSer-----ValArg	763
Qy	1543	GCTGTGTATGAGGGGAGCTGCTGAGCGCAT-----GGTGGCATT	1584
Db	764	PheValAlaGluArgly-----LysHisCysIleLeuAspValSerGlyAsnAla	780
Qy	1585	GTGAAGAGG-----	1593
Db	781	IleLysArgLeuGlnValAlaGlnLeuTyrProIleAlaIlePheIleLysProArgSer	800
Qy	1594	---GACGAGATCATGCCATCAACGCCAAGATTGTGACAGACTACACCCCTGGCTGAGGCT	1650

Db 801 LeuGluSerLeuMetGluMetAsnLysArgLeuThrGluGluGlnAlaLysLysThrTyr 820

QY 1651 GACGTCGCGCCAGAGAGCCCTGCAATCAGCGGGGGGACTGGATCGACCTTGCGTT 1707
 ||| ||||| ||| ::| ||||| ::||| ::|||

Db 821 AsparGAlaIle--LysLeuGluGlnGluPheGlycylTyrPheThrAlaIleVal 838

RESULT 15

DLG2_RAT

ID	DLG2_RAT	STANDARD;	PRT;	852 AA.
AC	Q63622;	Q62939; P70548;		
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Channel associated protein of synapse-110 (Chapsyn-110) (Synaptic density protein PSD-93) (Disks, large homolog 2).			
GN	DLG2.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxId=10116;			
[1]				
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=96310881; PubMed=8755482;			
RX	Kim E., Cho K.O., Rothschild A., Sheng M.;			
RA	"Heteromultimerization and NMDA receptor-clustering activity of			
RT	Chapsyn-110, a member of the PSD-95 family of proteins.";			
RL	Neuron 17:103-113(1996).			
[2]				
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=96193770; PubMed=8625413;			
RX	Brennan J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E.,			
RA	Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F.,			
RA	Froehner S.C., Bretl D.S.;			
RT	"Interaction of nitric oxide synthase with the postsynaptic density			
RL	protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";			
RL	Cell 84:757-767(1996).			
[3]				
RP	SEQUENCE FROM N.A.			
RP	Irie M., Hata Y., Takai Y.;			
RA	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
RL	-1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR			
CC	SUBUNITS AS WELL AS POTASSIUM CHANNELS.			
CC	-1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.			
CC	-1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-			
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstat-			
CC	ion of the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no			
CC	modified and this statement is not removed. Usage by and for commec-			
CC	entities requires a license agreement (See http://www.isb-sib.ch/anno			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; U49049; AAB53243.1; -			
CC	EMBL; U50717; AAC52643.1; -			
DR	EMBL; U53368; AAB48562.1; -			
DR	HSPSP; Q12959; 1PDR.			
DR	InterPro; IPR000619; Guanylate_kin.			
DR	InterPro; IPR001478; PDZ.			
DR	InterPro; IPR001452; SH3.			
DR	Pfam; PF00018; SH3; 1.			
DR	Pfam; PF00595; PDZ; 3.			
DR	Pfam; PF00625; Guanylate_kin; 1.			
DR	ProDom; PD000066; SH3; 1.			
DR	SMART; SM00072; GuKc; 1.			
DR	SMART; SM00228; PDZ; 3.			
DR	SMART; SM00326; SH3; 1.			
DR	PROSITE; PS00856; GUANYLATE_KINASE_1; 1.			
DR	PROSITE; PS50052; GUANYLATE_KINASE_2; 1.			
DR	PROSITE; PS50106; PDZ; 3.			
DR	PROSITE; PS50002; SH3; 1.			

KW SH3 domain; Repeat. 184 PDZ 1. 184
 FT DOMAIN 98 PDZ 2. 279
 FT DOMAIN 193 PDZ 3. 501
 FT DOMAIN 421 SH3. 606
 FT DOMAIN 536 GUANYLATE KINASE.
 FT DOMAIN 662 181 182 VR -> IL (IN REF. 2).
 FT CONFLICT 181 182 I -> M (IN REF. 2).
 FT CONFLICT 228 228 R -> K (IN REF. 2).
 FT CONFLICT 326 326 D -> E (IN REF. 3).
 FT CONFLICT 339 339 MISSING (IN REF. 2).
 FT CONFLICT 450 454 GD -> RK (IN REF. 2).
 FT CONFLICT 464 465 D -> H (IN REF. 2).
 FT CONFLICT 474 474 R -> P (IN REF. 2).
 FT CONFLICT 476 476 A -> D (IN REF. 2).
 FT CONFLICT 478 478 AAA -> LP (IN REF. 2).
 FT CONFLICT 484 486 A -> S (IN REF. 2).
 FT CONFLICT 506 506 H -> N (IN REF. 2).
 FT CONFLICT 569 569 L -> Q (IN REF. 2).
 FT CONFLICT 586 641 MISSING (IN REF. 2).
 FT CONFLICT 626 641 K -> A (IN REF. 3).
 FT CONFLICT 639 639 F -> L (IN REF. 1).
 FT CONFLICT 726 726 N -> Y (IN REF. 2).
 FT CONFLICT 733 733 E -> V (IN REF. 1).
 FT CONFLICT 749 749 L -> H (IN REF. 2).
 FT CONFLICT 756 756 KR -> NG (IN REF. 2).
 FT CONFLICT 791 794 T -> M (IN REF. 2).
 FT CONFLICT 794 794 F8D414A8B9CF5B09 CRC64;
 SQ SEQUENCE 852 AA; 94934 MW; 852

Alignment Scores:
 Pred. No.: 2.63e-05 Length: 852
 Score: 229.00 Matches: 145
 Percent Similarity: 36.09% Conservative: 108
 Best Local Similarity: 20.68% Mismatches: 226
 Query Match: 5.56% Indels: 222
 DB: 1 Gaps: 33

US-09-502-945-4 (1-2236) x DLG2_RAT (1-852)
 QY 208 GAGTGGCCGCTCTGCTGGAGACCTGAAGCTGGTGCATCAATGAA-----CCGAGCGCT 261
 DB 141 AspGlyArgLeuArgValAsnAspCysIleLeuArgValAsnGluValAspValSerGlu 160
 QY 262 CTGCTCTCTTTGATGCCATTCGGCGCTGATCCCACTGAAGCACCAGCTGGAATATGAT 321
 DB 161 ValSerHisSerLysAlaValGlu-----AlaLeuLysLeuAlaGlySerIleVal 177
 QY 322 CAGCTGACCCCGCGCTCCAGGAAGCTG-----AAGGAGTGGCTCTGGAC 369
 DB 178 ArgLeuTyrValArgArgArgProIleLeuGluThrValValGluIleLysLeuPhe 197
 QY 370 CGTCTGACCCGAGGCTCGGCTCGAGTGTGCGTGGTGGC-----CTG 414
 DB 198 Lys---GlyProLysGlyLeuGlyPheSerIleAlaGlyGlyValGlyAsnGlnHisIle 216
 QY 415 GAGTTTGGTGGCTTTCATCTCCACCTCATCAAGCGGTCAGCAGCAGCGCTC 474
 DB 217 ProGlyAspAsnSerIleTyrValThrLysIleAlaGlyGlyAlaAlaGlnLysAsp 236
 QY 475 GGG---CTCCAGTAGGGGAGAGATCGCCGATCAATGGATATTCATCTCTCTCT 531
 DB 237 GlyArgLeuGluValGlyArgLeuLeuMetValAsnAsnTyrSerLeuGluGluVal 256
 QY 532 ACCATGAGGAGTGCATCACTCATTCGA---ACCAAGAAACTGTGCTCCATCAAGTG 588
 DB 257 ThrHisGluGluAlaAlaIleLeuLysAsnThrSerAspValValTyrLeuLysVal 276
 QY 588 ----- 588
 DB 277 GlyLysProThrThrIleTyrMetThrAspProTyrGlyProProAspIleThrHisSer 296
 QY 589 -----AGACACATC-----GGCTGATGCCCGTGAA 615
 DB 589 -----AGACACATC-----GGCTGATGCCCGTGAA 615

297 TyrSerProProMetGluAsnHisLeuLeuSerGlyAsnAsnGlyThrLeuGluTyrLys 316
 616 AGCTCT----- 621
 317 ThrSerLeuProProIleSerProGlyArgTyrSerProIleProLysHisMetLeuVal 336
 622 -----CCTGATGAGCCCTCACTTGGCAGTATGTGGATCATGTTTGTG 663
 337 GluAspAspTyrThrArgProProGluProVal---TyrSerThrValAsnLysLeuCys 355
 664 TCGAATCTGGGGCGTCCGA----- 684
 356 AsLysProAlaSerProArgHisTyrSerProValGluCysAspLysSerPheLeuLeu 375
 685 -----GGC 687
 376 SerThrProTyrProHisTyrHisLeuGlyLeuLeuProAspSerAspMetThrSerHis 395
 688 AGCTGGGCTCCCTCGAATCGG----- 711
 396 SerGlnHisSerThrAlaThrArgGlnProSerValThrLeuGlnArgAlaIleSerLeu 415
 712 GAAACAAGGAGAAGAGTCTTCATCAGCTGCCGAGCTCCGAGCTTGGCTGCAGC 771
 416 GluGlyGluProArgLysValLeuHis---LysGlySerThrGlyLeuGlyPheAsn 434
 772 ATTTCCAGCGCCCTCCAGAGCTGTCATCTTTATCATCCCATGTG-----AAA 822
 435 IleValGlyGlyGlu---AspGlyGlyIlePheValSerPheIleLeuAlaGlyGly 453
 823 CTGCTCCCTCTGCTGAGTGGGATGGATGGGACCATGTTTCATGACAGCGGAG 1002
 454 ProAlaAspLeuSerGlyGlu-----LeuGlnArgGlyAspGlnIleLeuSerValasn 510
 883 GCGCTCGACTTCTTAACCTGATCACAAGAGCTGTAATGTCTGAAATAGCGCG 942
 472 GlyIleAspLeuArgGlyAlaSerHisGlnAlaAlaAlaLeuLysGlyAla--- 490
 943 AGCTGACCATCTCCATTTAGCTGCAGCTGGCGGAGCTGTTTCATGACAGCGGAG 1002
 491 GlyGlnThrValThrIleAlaGlnTyrGlnProGluAspTyrAlaArgPheGluAla 510
 1003 CGGCTGGCAGAGCGCGCAG----- 1023
 511 LysIleHisAspLeuArgGluGlnMetMetAsnHisSerMetSerSerGlySer 530
 1024 ---CCTGAGCTGCAGCGCAGAGCTTCTCATGCAAGAGCGCTGCGCATGAGTCCAAC 1080
 531 LeuArgThrAsnGlnLysArgSerLeuTyrValArgAlaMetPheAspTyrAspLysSer 550
 1081 AAG-----ATC 1086
 551 LysAspSerGlyLeuProSerGlnGlyLeuSerPheLysTyrGlyAspIleLeuHisVal 570
 1087 CTCAGGAGCAGCAGCAGATGGCGGCAAGAGGAGAAAGAAAT----- 1131
 571 IleAsnAlaSerAspGluTyrProGlnAlaArgValIleLeuAspGlyAspSer 590
 1132 -----GCCAGAGGAGCAGCAGAGAGAAATGAGATACCGGAG 1170
 591 GluGluMetGlyValIleProSerLysArgValGluArgLysGluArgAlaArgLeu 610
 1171 GAGATGGAGCAG-----ATTGTAGAGGAGGAGAGAGAGTTTAAAGAG 1212
 611 LysThrValLysPheAsnAlaLysProGlyValIleAspSerLysGlyAspIleProGly 630
 1213 CAATGGGAGAGAGCTGGGCTCAAG-----GAACGCTACTCTTGCT 1257
 631 LeuGlyAspGlyTyrGlyThrLysThrLeuArgGlyGlnGluAspLeuLeuSer 650
 1258 ---AAACCATCATCTCTGAG-----GTACACCCAGTACCTCCCTCGCAAGCA 1302
 651 TyrGluProValThrArgGlnGlnIleAsnTyrThrArgProValIleLeuGlyPro 670

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:20:42 ; Search time 76.6811 Seconds

(without alignments)

12016.549 Million cell updates/sec

Title: US-09-502-945-4

Perfect score: 4120

Sequence: 1 cctgcccgcgcgcgtgcg.....tattttccagcttaaaaaa 2236

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Xgapop 10.0 , Xgapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p.model -DEV=xlx
-Q/cgn2_1/USPRO_pool/US09502945/runat_14032003_101058_19113/app_query.fasta_1.10979
-DB-SPTREMBL_21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_565 -runat_14032003_101058_19113 -NCPU=6 -ICPU=3
-NO_XLPAY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2720.5	66.0	533	4 Q96B29	Q96B29 homo sapien

2	2674	64.9	548	11	Q9ES65
3	2670	64.8	548	11	Q91XD1
4	2459	59.7	910	11	Q9ES64
5	548	13.3	112	11	Q9CVG7
6	425.5	10.3	493	5	Q9W443
7	407.5	9.9	963	4	Q9P202
8	406.5	9.9	906	4	Q8WVS2
9	283.5	6.9	517	4	Q9H5P4
10	275	6.7	904	11	Q921G9
11	272	6.6	202	4	Q9NTP3
12	269.5	6.5	2042	4	Q75970
13	269	6.5	2054	11	Q55164
14	264.5	6.4	835	4	Q9EC69
15	263.5	6.4	1564	4	Q14160
16	262.5	6.4	1630	4	Q8WVW8
17	262	6.4	2055	11	Q8VBX6
18	262	6.4	2055	11	Q8VBV5
19	260.5	6.3	1665	11	Q8V111
20	260	6.3	2055	11	Q8VBY0
21	259.5	6.3	582	4	Q8WX31
22	259.5	6.3	721	11	Q91WJ1
23	259	6.3	453	4	Q43798
24	259	6.3	2055	11	Q921K3
25	256.5	6.2	1524	4	Q15249
26	256.5	6.2	1552	4	Q60833
27	256.5	6.2	1582	4	Q43742
28	256	6.2	526	11	Q08783
29	254.5	6.2	944	11	Q922S3
30	252.5	6.1	927	11	Q62402
31	251	6.1	1769	6	Q97758
32	243.5	5.9	345	4	Q96M29
33	241.5	5.9	1012	5	Q21075
34	238.5	5.8	1134	4	Q9H3N9
35	231	5.6	1756	5	Q9VBE4
36	230	5.6	1465	5	Q8T003
37	229	5.6	852	11	Q91XM9
38	227.5	5.5	967	5	Q9B179
39	227.5	5.5	1064	5	Q18165
40	221	5.4	871	5	Q9NB04
41	220	5.3	871	5	Q9X235
42	220	5.3	2460	11	Q64512
43	215.5	5.2	1462	4	Q96QZ7
44	214.5	5.2	1455	4	Q60510
45	214	5.2	974	5	Q960N4

ALIGNMENTS

RESULT 1

Q96E29 ID Q96B29 PRELIMINARY; PRT; 533 AA.
AC Q96B29
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to PDZ-73 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL: BC016057; AAH16057.1;
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 3.
DR PROSITE: PS0106; PDZ; 3.
SQ SEQUENCE 533 AA; 60314 MW; A9D64A3B011E2FBD CRC64;

Alignment Scores:

Pred. No.: 1.4e-179 Length: 533

Score: 2720.50 Matches: 532
Percent Similarity: 96.55% Conservative: 0
Best Local Similarity: 96.55% Mismatches: 0
Query Match: 66.03% Indels: 19
DB: 4 Gaps: 1

US-09-502-945-4 (1-2236) x Q96B29 (1-533)

```
QY 97 ATGGACCAAGAGTGGCCGCGAGAAATTCGGGATAGGTGGATTTCTGATGAAAATGAT 156
DB 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
QY 157 GCAGAGAGGACTATCTCTATGATGTGTCGGAATGTACACAGACCATGACGTGCC 216
DB 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
QY 217 GNGCTCGTGGGACCTGAAGCTGTCATCAATGAACCCAGCCGCTGCTCTGTTGAT 276
DB 41 ValIleValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
QY 277 GCATTGCGCCGCTGATCCACTGAAGCACCCAGGTGGAATATGATCAGCTGACCCCGG 336
DB 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
QY 337 CGCTCCAGGAAGTGAAGAGGTGCTGACCGTCTGACCCCGAGAGGCTCGGCCCTG 396
DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
QY 397 ACTGTCGTGGTGGCTGAGTGTGGCTGGCTCTCATCTCCACCTCATCAAGGC 456
DB 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleGly 120
QY 457 GGTGAGCAGACAGCTCGGGCTCCAGGTAGGGAGCAGATCTCGGATCAATGGAT 516
DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
QY 517 TCCATCTCTCTGTACCATCAGGAGGTCTCAACCTCATTCGAACCAAGAAACTGTG 576
DB 141 SerIleSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
QY 577 TCCATCAAGTGAACATCGCCCTGATCCCGTGAAGCTCTCCATGACGCCCTC 636
DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
QY 637 ACTTGCCAGTATGTGATCAGATTTGTCTGGAATCTGGGGGTGCGAGGACCTGGGC 696
DB 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200
QY 697 TCCCTCGAAATCGGAAACAGAGAGAGAGGTCTTCATCAGCTGTGAGCTCCCGA 756
DB 201 SerProGlyAsnArgGluAsnLysGluLysLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCCTTGGCTGACGATTTCCAGCGGCCCATCCAGAGCTGGCATCTTATCAGCCAT 816
DB 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTCAACCTGGCTGCTGCTGCTGAGTGGGATGAGATAGGGACAGATGTCGAA 876
DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCAATGGCTGCTGCTGCTGCTGATCAACAGAGGCTGTAATGCTGTAATAAT 936
DB 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGlu----- 273
QY 937 AGCGCAGCCTGACCATCTCTCCATGTAGTCGAGCTGGCGGAGCTGTTTCATGACAG 996
DB 274 -----GlyArgGluLeuPheMetThrAsp 281
QY 997 CGGAGCGGCTGGCAGAGCGCGGCGGCTGAGCTGCGAGCGGAGGCTCTCATGCGAG 1056
DB 282 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 301
QY 1057 AAGCGGCTGGCGATGGAGTCCAAACAGATCTCCAGGAGCAGGAGATGGAGCGGCAA 1116
```

```
DB 302 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln 321
QY 1117 AGGAGAAAGAAATTCGCCAGAGGAGGAGAGAGAAATGAGAGATACCCGAGAGAGATG 1176
DB 322 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 341
QY 1177 GAACAGATTGTAGAGGAGGAGAGAGAGTTTAAAGCAATGGGAAGAGACTGGGGCTCA 1236
DB 342 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 361
QY 1237 AAGGAACAGCTACTCTTGTCCCTAAAACCATCCTCCTCAGGTACACCCAGTACCCCTTCGC 1296
DB 362 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 381
QY 1297 AAGCAAAAGTATGATCAGGAGTGGAAACCTGAGCTCGAGCCCGCAGATGACCTGGATGA 1356
DB 382 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 401
QY 1357 GGCAGGAGGAGGAGGAGGAGGAGGATTCGGGAAATATGAGGAAGCTTTGAGCCCTAC 1416
DB 402 GlyThrGluGluGlnGlyGluGlnAspPheArgLysTyrGluGluGlyPheAspProTyr 421
QY 1417 TCTATGTTCAACCCAGAGCAGATCATGGGAAGGATGTCCGGCTCCTACGCTCAAGAAAG 1476
DB 422 SerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLys 441
QY 1477 GAGGATCCTTAGACCTGGCCCTGGAAGCGGTGTGAGCTCCCTCCATGGGAAGGTGGTC 1536
DB 442 GluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProIleGlyLysValVal 461
QY 1537 GTTCTGCTGTGTATGAGCGGGGAGCTGCTGAGCGGATGTGTCATTTGAAAGGGGAC 1596
DB 462 ValSerAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAsp 481
QY 1597 GAGATCATGGCAATCAAGCGCAAGATTGTGACACACTACACCTGGCTGAGGCTGACGCT 1656
DB 482 GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAla 501
QY 1657 GCGCTGCAAGAGGCTGGAATCAGGGCGGAGCTGATGATGACCTGATGTTGGCTGCTGCG 1716
DB 502 AlaLeuGlnLysAlaTrpAsnGlnGlyLysAspTrpIleAspLeuValValAlaValCys 521
QY 1717 CCCCAGAGGAGTATGACCATGACGTGACCTTC 1749
DB 522 ProProLysGluTyrAspAspGluLeuThrPhe 532
```

RESULT 2

```
Q9ES65 PRELIMINARY; PRT; 548 AA.
AC Q9ES65;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Harmonin isoform a1.
GN 2010016F01RIK OR USH1C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20428180; PubMed=10973247;
RA Verpy E., Leibovici M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,
RA Mansour A., Blanchard S., Kobayashi I., Keats B.J., Slim R., Petit C.;
RT "A defect in harmonin, a PDZ domain-containing protein expressed in
RL the inner ear sensory hair cells, underlies Usher syndrome type 1C.";
RL Nat. Genet. 26:51-55(2000).
DR EMBL; AF228924; AAG12457.1;
DR HSSP; P29476; 10AV.
DR MGD; MGI:1919338; 2010016F01RIK.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
```

DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS50106; PDZ; 3.
SQ SEQUENCE 548 AA; 61712 MW; 44A245AA32091070 CRC64;

Alignment Scores:

Pred. No.:	2,31e-176	Length:	548
Score:	2674.00	Matches:	521
Percent Similarity:	96.73%	Conservative:	12
Best Local Similarity:	94.56%	Mismatches:	14
Query Match:	64.90%	Indels:	64
DB:	11	Gaps:	1

US-09-502-945-4 (1-2236) x Q9ES65 (1-548)

QY	97	ATGACCGAAAGTGGCCCGAGAAATCCGGCATAAAGTGGATTTCCTGATTGAAATGAT	156
Db	1	MetAspArgLysValAlaAArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp	20
QY	157	GCAGAGAAGACTATCTCTATGATGCTCGGAATGTACACAGACCATGGACGTGGCC	216
Db	21	AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla	40
QY	217	GTGCTCTGGGAGACCTGAAGCTGGTCAATCAATCAACCCAGCCGCTGCTGCTGTTGAT	276
Db	41	ValLeuValGlyAspLeuLysLeuValIleAsnGluProAsnArgLeuProLeuPheAsp	60
QY	277	GCCATTCCGGCGCTGATPCCCACTCAAGCACACAGGTGGAATATATCATGAGTGCACCCCGG	336
Db	61	AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg	80
QY	337	CGCTCCAGGAAGCTGAAGAGGTGCGTCTGGACCGTCTGCACCCGAAAGCCCTCGGCCTG	396
Db	81	ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu	100
QY	397	AGTGTGGTGTGGCGCTGGAGTTGGCTGTGGGCTCTTCACTCCACCTCATCAAGGC	456
Db	101	SerValArgGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly	120
QY	457	GGTCAGGCAGACAGCGTCGGCTCCAGGTAGGGACGAGATCTCCGGATCAATGGATAT	516
Db	121	GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr	140
QY	517	TCCATCTCTCTGTACCCATGAGGAGGTCAATCAACTTCGAACCAAGAAACTGTG	576
Db	141	SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal	160
QY	577	TCCATCAAGTGACACATCGCCCTGATCCCGCTGAAAGCTCTCCTCATGAGCCCTC	636
Db	161	SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProGluGluSerLeu	180
QY	637	ACTTGGCAGTATGTGGATCAGTTGTTCGGAATCTGGGGGGCTGCGAGGCAGCCTGGCC	696
Db	181	LysTrpGlnTyrValAspGlnPheValSerGluSerGlyClyValArgGlyLeuGly	200
QY	697	TCCCTCGAATCGGNAACACAGGAGAAGGTCTTCATCAGCCTGTGAGTCCCGA	756
Db	201	SerProGlyAsnArgThrLysGluLysLysValPheIleSerLeuValGlySerArg	220
QY	757	GGCCTTGGCTGCAGCATTTCCACGGCCCATCCAGAGCCTGGCATCTTTATCAGCCAT	816
Db	221	GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheValSerHis	240
QY	817	GTAAACCTGGCTCCCTGTCTGCTGAGGTGGGATTTGGAGATAGGGGACACAGATTTCGAA	876
Db	241	ValLysProGlySerLeuSerAlaGluValGlyLeuGluThrGlyAspGlnIleValGlu	260
QY	877	GTCAATGGCTGCAGTCTCTTAACCTTGGATCACAGGAGGCTGTAAATGTGCTGAAAT	936
Db	261	ValAsnGlyIleAspPheThrAsnLeuAspHisLysGluAlaValAsnValLeuLysSer	280
QY	937	AGCGGCAGCTGACCATCTCCATTGTAGCTGCAGCTGGCGGAGCTGTTCATGACAGAC	996
Db	281	SerArgSerLeuThrIleSerIleValAlaGlyArgGlyLeuPheMetThrAsp	300

```
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 3.
DR PROSITE: PS0106; PDZ; 3.
SQ SEQUENCE 548 AA: 61711 MW: B00C4F00920ED76B CRC64:

Alignment Scores:
Pred. No.: 4,37e-176 Length: 548
Score: 2670.00 Matches: 520
Percent Similarity: 96.73% Conservative: 13
Best Local Similarity: 94.37% Mismatches: 14
Query Match: 64.81% Indels: 4
DB: 11 Gaps: 1

US-09-502-945-4 (1-2236) x Q91XD1 (1-548)
QY 97 ATGGACCGAAAGTGGCCGAGAAATCCGGCATAAAGTGGATTTCTGTATGAAATGAT 156
DB 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
QY 157 GCAGAGAAGACTATCTCTATGATGTGTCGGAATGTACACAGACCATGGACGTGGCC 216
DB 21 AlaGluLysAspPyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
QY 217 GTGCTGTGGGACACCTGAAGCTGGTGCATCAATGAACCCAGCCGCTGCTCTGTTGAT 276
DB 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProAsnArgLeuProLeuPheAsp 60
QY 277 GCCATTTCGGCCGCTGATCCACATGAAGCACACAGGTGGAATATGATCAGCTGACCCCGG 336
DB 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
QY 337 CGCTCCAGGAAGCTGAAGAGGTGCTGTGGACCGTCTGCACCCGGAAGCCCTCGCCCTG 396
DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
QY 397 AGTGTGCTGCTGCTGGCTGGATTTGGCTGTGGCTCTTCATCTCCACATCAATCAAGGC 456
DB 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
QY 457 GGTGAGCAGACAGCTGCGGCTCCAGGTAGGGAGGAGATCGTCGGATCAATGATGATAT 516
DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
QY 517 TCCATCTCTCTGTACCCATGAGGAGGTATCACTCACTTCGACCAAGAAACTGTG 576
DB 141 SerIleSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
QY 577 TCCATCAAACTGAGACATCGCCCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTC 636
DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProGluGluSerLeu 180
QY 637 ACTTGGCAGTATGTGATCAGTTTGTGTCGGAATCTGGGGCGTGGCAGCAGCCCTGGC 696
DB 181 LysTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlyGlyLeuGly 200
QY 697 TCCCTCGAATCGGAAACACAGGAGAAGAGTCTTCATCAGCCTGGTAGCCTCCCGA 756
DB 201 SerProGlyAsnArgThrThrLysGluLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCCTTGGCTGCAGATTTCCAGCGGCCCATCCAGAGCTGCGATCTTTATCAGCCAT 816
DB 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheValSerHis 240
QY 817 GTGAACCTGGCTCCCTGTCTGTGAGGTGGGATTTGAGATAGGGACCATGATTCGAA 876
DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluThrGlyAspGlnIleValGlu 260
QY 877 GTCATGGCTGCAGCTTCTTCACTCGATCACAAGGAGCTGTAATGTGCTGGAATAT 936
DB 261 ValAsnGlyIleAspPheThrAsnLeuAspHisLysGluAlaValAsnValLeuLysSer 280
QY 937 AGCCGAGCCTGACCATCTCCATGTAGTGCAGCTGGCGGAGCTGTTCATGACAGAC 996
DB 937 AGCCGAGCCTGACCATCTCCATGTAGTGCAGCTGGCGGAGCTGTTCATGACAGAC 996
```

```
Db 281 SerArgSerLeuThrIleSerIleValAlaGlyAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGAGCGCTGGCAGAGCGCGCAGCGTGAAGTGGAGCGCAGGAGGAGCTTCTCATGCAG 1056
DB 301 ArgGluArgLeuGluGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuMetGln 320
QY 1057 AAGCGCTGGCGATGAGTCCCAACAAGATCTCCAGGAGCAGCAGGAGATGAGCGGCAA 1116
DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnGlnGlnGlnGln 340
QY 1117 AGGAGAAAGAAATTCGCCAGAGCAGCAGAGGAAATAGAGATACCCGAAAGGAGATG 1176
DB 341 ArgArgLysGluIleAlaGlnLysAlaGluGluAsnGluArgTyrArgLysGluMet 360
QY 1177 GAACAGATTGTAGAGGAGGAAGATTTAAGAACAAATGGAGAGAGACTGGGGCTCA 1236
DB 361 GluGlnIleSerGluGluGluLysPheLysLysGlnTrpLysGluAspTrpGlySer 380
QY 1237 AAGGAGACGACTCTTCCTTAAACCATCATCTGCTGAGGTACACCCAGTACCCCTGCGC 1296
DB 381 LysGluGlnLeuIleLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCCAAAGTATGATCAGGAGTGGAACTGAGCTCGAGCCGCGAGATGACTGTGATGA 1356
DB 401 LysProLysTyrAspGlnGlyVal-----GluProAlaAspHisLeuAspGly 416
QY 1357 GGCAGGAGGAGCAGGAGCAGGAGTTCGCGAAATATGAGGAAGGCTTTGACCCCTAC 1416
DB 417 SerThrGluGluGlnArgGlnAspPheArgLysTyrGluGluGlyPheAspProTyr 436
QY 1417 TCTATGTTTCAACCCAGCAGCAGATCATGGGAGGATGTCCGGCTCCTACGATCAAGAAG 1476
DB 437 SerMetPheSerProGluGlnIleAlaGlyLysAspValArgLeuLeuArgIleLysLys 456
QY 1477 GAGGATCTTATAGACTGCGCCCTGGAAGCGGTGGACTCCCCATTTGGGAAGGTGTC 1536
DB 457 GluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProValGlyLysValVal 476
QY 1537 GTTCTGTGTGTATGAGCGGGAGCTGCTGAGCGGCATGTGGCATTTGTGAAAGGGAC 1596
DB 477 ValSerAlaValTyrGluGlyAlaAlaGluArgHisGlyGlyValValLysGlyAsp 496
QY 1597 GAGATCATGGCAATCAAGCGCAAGATTTGTGACAGACTACACCTGGCTGAGGCTCACGCT 1656
DB 497 GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaGlu 516
QY 1657 GCCTTGCAGAGCCCTGGAATCAGGCGGGAGTGGACTGGATCGACTTGTGTGCTGCTGC 1716
DB 517 AlaLeuGlnLysAlaTrpAsnGlnGlyGlyAspTrpIleAspLeuValAlaValCys 536
QY 1717 CCCCAGAGGAGTATGACGATGAGCTGACCTTC 1749
DB 537 ProProLysGluTyrAspAspGluLeuThrPhe 547

RESULT 4
Q9ES64 PRELIMINARY: PRT; 910 AA.
AC Q9ES64;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Harmonin isoform b3.
GN 2010016F01RIK OR USHC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20428180; PubMed=10973247;
RA Verpy E., Leibovici M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,
RA Mansour A., Blanchard S., Kobayashi I., Keats B.J., Slim R., Petit C.;
RT "A defect in harmonin, a PDZ domain-containing protein expressed in
```

the inner ear sensory hair cells, underlies Usher syndrome type 1C.;

RL Nat. Genet. 26:51-55(2000).

DR EMBL; AF228925; AAG12458.1; ..

DR HSP; P29476; IQAV

DR MGD; MGI:1919338; 2010016P01R1k.

DR InterPro; IPR001478; PDZ.

DR InterPro; IPR002965; P_Rich_extensn.

DR Pfam; PF00595; PDZ; 3.

DR PRINTS; PR01217; PRICHEXTENS.

DR SMART; SM00228; PDZ; 3.

DR PROSITE; PS0106; PDZ; 3.

SQ SEQUENCE 910 AA; 102284 MW; 427B97953BA5D941 CRC64;

Alignment Scores:

Pred. No.: 2,04e-161 Length: 910

Score: 2459.00 Matches: 535

Percent Similarity: 59.62% Conservative: 23

Best Local Similarity: 57.16% Mismatches: 50

Query Match: 59.68% Indels: 328

DB: 11 Gaps: 10

US-09-502-945-4 (1-2236) x Q9ES64 (1-910)

QY 97 ATGGACCGAAGTGGCCGCGAGAAATCCGGCATAAGTGGATTTCTGATTGAAATGAT 156

Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20

QY 157 GCAGAGAGGACTACTCTATGATGCTGCGGAATGTACACAGACCATGAGCTGGCC 216

Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40

QY 217 GTGCTCGTGGGAGACTGAAGTGTGTCATCAATGAACCCAGCGTCCCTCTGTTGAT 276

Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProAsnArgLeuProLeuPheAsp 60

QY 277 GCCATTCCGGCGCTGATCCCACTGAAGCACCAGTGGGAATATGATCAGCTGACCCCGG 336

Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80

QY 337 CGCTCCAGAGAGCTGAAGAGGTGGCTGACCGCTGACCGCTGACCGCGAGGCGCTGGCG 396

Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100

QY 397 AGTGTGCTGTGGCTGGAGTTGGCTGTGGCTGCTTCATCTCCACCTCATCAAGGC 456

Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120

QY 457 GGTACGGCAGACAGCTCGGCTCCAGTAGGGAGACGATCGTCCGGATCAATGGATAT 516

Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140

QY 517 TCCATCTCTCTGTACCATGAGGAGGTGTCATCACTCATCTCGAACCAAGAACTGTG 576

Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160

QY 577 TCCATCAAGTGAGACACATCGGCTGATCCCGTGAAGCTCTCTGATGAGCCCTC 636

Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProGluGluSerLeu 180

QY 637 ACTTGGCAGTAGTGGATTCAGTTTGTGCGGAATCTGGGGCGTGGGAGGAGCTGGGC 696

Db 181 LysTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlyGlyLeuGly 200

QY 697 TCCCTTGGAAATCGGGAACCAAGAGAGAGTCTTTCATCAGCCTGTAGTGGCTCCCGA 756

Db 201 SerProGlyAsnArgThrThrLysGluLysLysValPheIleSerLeuValGlySerArg 220

QY 757 GGCCTTGGCTGACCATTTCCAGCGGCCCATCCAGAGGCTGGCATCTTTATCAGCAT 816

Db 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheValSerHis 240

QY 817 GTGAACCTGGCTCCCTGCTGCTGAGTGGGATTTGGAGATAGGGACCATGTCGAA 876

Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluThrGlyAspGlnIleValGlu 260

QY 877 GTCAATGGCGTCGACTTCTTAACCTGGATCACAAGAGGCTGPAATGTGCTGAAAT 936

Db 261 ValAsnGlyIleAspPheThrAsnLeuAspHisLysGluAlaValAsnValLeuLysSer 280

QY 937 AGCCGACGCTGACCATCTCATGTAGCTGCAGCTGCGGGAGCTGTTTCATCAGACAG 996

Db 281 SerArgSerLeuThrIleSerIleValAlaGlyArgGluLeuPheMetThrAsp 300

QY 997 CGGAGGCGGTGGCAGAGCGCGCAGCTGAGCTGCAGCGGAGGAGCTCTCATCAG 1056

Db 301 ArgGluArgLeuGluGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320

QY 1057 AAGCGGTGGGATGGAGTCCACAAGATCTCCAGAGCAGCAGAGATGGAGCGCAA 1116

Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnMetGluArgGln 340

QY 1117 AGGAGAAAAGAAATGCCAGAGGCGCAGAGGAAATGAGAGATACCGGAGGAGATG 1176

Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360

QY 1177 GAACAGATTCTAGAGGAGGAGAGAGTAAAGCAATGGGAAGACTGGGGCTCA 1236

Db 361 GluGlnIleSerGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380

QY 1237 AAGAACAGCTACTTTCCTTAAACCATCAGCTGCTGAGTACACCCAGTACCCCTTCG 1296

Db 381 LysGluGlnLeuIleLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400

QY 1297 AAGCCA----- 1302

Db 401 LysProLysSerPheGlyTrpPheTyrArgTyrAspGlyLysPheProThrIleArgLys 420

QY 1303 -----AAGTATGAT----- 1311

Db 421 LysAlaLysGluLysLysLysAlaLysTyrAspSerLeuGlnAspLeuArgLysAsnLys 440

QY 1311 ----- 1311

Db 441 LysGluLeuGluPheGluGlnLysLeuTyrLysGluLysGluMetLeuGluLysGlu 460

QY 1311 ----- 1311

Db 461 LysGlnLeuLysIleAsnArgLeuAlaGlnGluValSerGluThrGluArgGluAspLeu 480

QY 1311 ----- 1311

Db 481 GluGluSerGluLysThrGlnTyrTrpValGluArgLeuCysGlnThrArgLeuGluGln 500

QY 1312 -----CAGGAGTGGAACTGACCTC----- 1332

Db 501 IleSerSerAlaGluAsnGluIleProGluMetThrThrGlyProProProProPro 520

QY 1332 ----- 1332

Db 521 SerValSerProLeuAlaProProLeuArgArgPheAlaGlyGlyIleHisLeuHisThr 540

QY 1332 ----- 1332

Db 541 ThrAspLeuAspIleProLeuAspMetPheTyrTyrProLysThrProSerAla 560

QY 1332 ----- 1332

Db 561 LeuProValMetProHisProProSerValAsnSerProSerLysValProAlaProPro 580

QY 1332 ----- 1332

Db 581 ValLeuProSerSerGlyHisValSerSerSerSerProTrpValGlnArgThrPro 600

QY 1332 ----- 1332

Db 601 ProProIleProLeuProProProProSerIleProThrGlnAspLeuThrProThrArg 620

```
QY 1332 ----- 1332
Db 621 ProLeuProSerAlaLeuGluGluAlaLeuGlyAsnHisProPheArgThrGlyAspPro 640
QY 1333 ---GAGCCGACAGATGACCTGGATGGAGCAGC----- 1362
Db 641 GlyHisProAlaAspTrpGluAlaAsnThrHisSerGlyLysProSerSerSerPro 660
QY 1362 ----- 1362
Db 661 ThrThrGluArgSerPheProAlaProAlaProLysThrPheCysProSerProGlnProPro 680
QY 1362 ----- 1362
Db 681 ArgGlyProGlyValSerThrIleSerLysProValMetValHisGlnGluHisAsnPhe 700
QY 1362 ----- 1362
Db 701 ValTyrArgProAlaValLysSerGluValLeuProGlnGluMetLeuLysArgMetVal 720
QY 1363 -----GAGGAGCAGGAGCAGAGATTCGGAATATATGAGGAAGCTTTCACCCCTAC 1416
Db 721 ValTyrGlnThrAlaPheArgGlnAspPheArgLysTyrGluGluGlyPheAspProTyr 740
QY 1417 TCTATGTTCCACCCAGCAGATCATGGGGAAGGATGTCCTGCTAGCATCAAGAG 1476
Db 741 SerMetPheSerProGlnIleAlaGlyLysAspValArgLeuLeuArgIleLysLys 760
QY 1477 GAGGATCTTACCTGACCTGCGCCCTGGAAGCGGTGTGGACTCCCCCATTTGGGAAGTGGPC 1536
Db 761 GluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProValGlyLysValVal 780
QY 1537 GTTTCGTGCTGTATGAGCGGGAGCTGCTGAGCGGCATGTGGCATTTGTGAAGGGAC 1596
Db 781 ValSerAlaValTyrGluGlyAlaAlaGluArgHisGlyValValLysGlyAsp 800
QY 1597 GAGATCATGCAATCAACGCAAGATTGTGACAGACTACACCTGCTGAGGCTGACGCT 1656
Db 801 GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAla 820
QY 1657 GCCTTCGAGAGCGCCATGCAATCAGCGGGGACTGATGACCTTGTGTGCGGCTGC 1716
Db 821 AlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValValAlaValCys 840
QY 1717 CCCCAGAGATCATCAGATGAGCTGACCTTCTGCTGAGTCCAAAGGGGAAACCAA 1776
Db 841 ProProLysGluTyrAspAspGluLeuSerSerLeuProSerSerAlaAlaGluSer 859
QY 1777 ATTCACGCGTTAGGAACAGTGTGACTCCGCGCCACCTCTGTGAACACAAAGCCTCGGACC 1836
Db 860 -----ProGlnLeuAla-----ArgLys 865
QY 1837 AGCCTTGAGAGAGCCACATGACACACACAGATGCGATCTTGGACCTGAATCATCA 1896
Db 866 GlnLeuGlu---AlaTyrGluProValCysArg---HisGlyPhePheLeuGlnLeuGlu 883
QY 1897 CCGAGGAATCTCAACTCCCTTTGGCCCTGAACACAGGGCCAGATAGAGAACACCTCGGC 1956
Db 884 ProThrAsnLeuLeuLeu-----LysSerArgGluArgAsnGlnThr--- 897
QY 1957 CACTTTTTTGAAGGCCAATGTGGAGGAAGGGAGCAGCCGCGC 2000
Db 898 -----Asp--ProSerTrpArgProAlaSerSerAlaPro 908
RESULT 5
Q9CVG7
ID Q9CVG7 PRELIMINARY; PRT; 112 AA.
AC Q9CVG7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 2010016F01RIK protein (Fragment).
```

```
GN 2010016F01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK008274; BAB25568.1; -.
DR HSSP; Q12923; 3PDZ.
DR MGD; MGI:1919338; 2010016F01RIK.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER 1
SQ SEQUENCE 112 AA; 11965 MW; BB42F80C7B3AB196 CRC64;
Alignment Scores:
Pred. No.: 1,01e-29 Length: 112
Score: 548.00 Matches: 105
Percent Similarity: 98.20% Conservative: 4
Best Local Similarity: 94.59% Mismatches: 2
Query Match: 13.30% Indels: 0
Db: 11 Gaps: 0
US-09-502-945-4 (1-2336) x Q9CVG7 (1-112)
QY 1417 TCTATGTTACCCAGACAGATCATGGGAGAGGATGTCCGGCTCTACGATCAAGAAG 1476
Db 1 SerMetPheSerProGluGlnIleAlaGlyLysAspValArgLeuLeuArgIleLysLys 20
QY 1477 GAGGATCCTTAGACTGGCCCTGGAAGCGGTGTGGACTCCCCATTTGGGAAGGTGTC 1536
Db 21 GluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProValGlyLysValVal 40
QY 1537 GTTTCGTGCTGTATGAGCGGGAGCTGCTGAGCGGCATGTGTGGCATTTGTCAAAGGGAC 1596
Db 41 ValSerAlaValTyrGluGlyAlaAlaGluArgHisGlyGlyValValLysGlyAsp 60
QY 1597 GAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTCGCTGAGGCTACCGCT 1656
Db 61 GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaGlu 80
QY 1657 GCCCTGACAGAGCCCTGGAAATCAGGGCGGGGACTGGATCGACCTTGTGGTTCCTCTGC 1716
Db 81 AlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValValAlaValCys 100
QY 1717 CCCCAGAGGAGTATGACGATGAGCTGACCTGACCTTC 1749
Db 101 ProProLysGluTyrAspAspGluLeuThrPhe 111
```



```
Db 494 GluHisAlaThrMetAlaTyrTyrLeuAspGlu-:::||||| |||||||
Qy 1168 -----AAGGAGTGGACAGATTGTAGAGGAGAGAGAAAGTTTAAAGACAA 1215
Db 513 GluAlaLeuValMetAlaLeuPheTyrLeuLeuAsnThrHisAlaLysPheSerLeuLeu 532
Qy 1216 TGGGAA-----GAAGACTGGGGCTCAAGGAACAGACTACTCTTG 1254
Db 533 SerGluValArgGlyThrIleSerProGlnAspLeuGluArgPheAspHisLeuValLeu 552
Qy 1255 CCTAAACCATCAGTCTGAGGTACACCCAGTACCCCTTCGGAAGCCA----- 1302
Db 553 ArgArg-----GluIleGluSerMetLysAlaArgGlnProProGlyProGly 568
Qy 1303 -----AAGTAT-----GATCAG 1314
Db 569 AlaGlyAspThrTyrSerMetValSerTyrSerAspThrGlySerSerThrGlySerHis 588
Qy 1315 GGAGTGGAACTGAGCTCGAGCCGCA---GATGACCTGATGGAGGCACGAGGAGCAG 1371
Db 589 GlyThrSerThrValSerSerAlaArgAsnThrLeuAsp-----LeuGluGluThr 606
Qy 1372 GGAGAG-----CAGGATTTCGGG 1389
Db 607 GlyGluAlaValGlnGlyAsnIleAsnAlaLeuProAspValSerValAspAspValArg 626
Qy 1390 AAATATGAGGAGGCTTACCCCTACTCTATGTTTC-----ACCCAGAGCAGATCATG 1443
Db 627 SerThrSerGlnGlyLeuSerSerPheLysProLeuProArgProProLeuAlaGln 646
Qy 1444 GGAAGAGTGTCCGGCTCTACGCATCAAGAAGAGGAGGATCTTACACTGGCCCTGGAA 1503
Db 647 GlyAsnAspLeuProLeuGlyGlnProArgLysLeuGlyArgGluAspLeuGlnProPro 666
Qy 1504 GCGGTGTGGACTCCCCCATTTGGGAGGTGGTCTGCTGTGTATGAGCGGGAGCT 1563
Db 667 SerSerMetProSerCysSerGly---ThrValPheSerAlaProGlnAsnArgSerPro 685
Qy 1564 GCTGAGCGCATGGTGGCATTTGTGAAGGGGAGGAGATCATGCGCAATCAACGGCAAGATT 1623
Db 686 Pro-AlaGly----- 688
Qy 1624 GTGACAGACTACACCTGGCTGAGGCTGACGCTGCCCTGCACAGAGCCTGGAATCAGGCG 1683
Db 689 -----ThrAlaPro-----ThrPr 693
Qy 1684 GGGAGTGGATGACCTTGTGTGTCGCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTG 1743
Db 693 oGlyThrSerSerAlaGlnAspLeuProSerSer-Pro----- 705
Qy 1744 ACTTCTTCTGAAGTCCAAAGGGGAAACCAATTCACGCCCTTAGGAACAGTGAAGCTC 1803
Db 706 ---IleTyrAlaSerValSerProAlaAsnProSerSerLys----- 718
Qy 1804 CGGCCCCACCTCGTGAACAAAGCCTCGACACGCTTGACAGAGGCCATGACACAC 1863
Db 719 -----ArgProLeuAspAlaH 724
Qy 1864 ACCAGATGGCATCTTGGGACCTGAATCTATCACCAGGAATCTCAAACTCCCTTTGGCC 1923
Db 724 is-----LeuAlaLeuValAsnGlnHisProIleGlyP 735
Qy 1924 CTGAACCA 1931
Db 735 roPhePro 737
RESULT 8
Q8WVS2
AC Q8WVS2 PRELIMINARY; PRT; 906 AA.
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
```

```
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 96.4 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARY;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014524; AAH14524.1; -.
DR InterPro; IP001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS0106; PDZ; 3.
KW Hypothetical protein.
SQ SEQUENCE 906 AA; 96428 MW; D8D75D87A41A5E05 CRC64;

Alignment Scores:
Pred. No.: 1.39e-19 Length: 906
Score: 406.50 Matches: 194
Percent Similarity: 37.57% Conservative: 93
Best Local Similarity: 25.39% Mismatches: 270
Query Match: 9.87% Indels: 207
DB: 4 Gaps: 28
```

```
US-09-502-945-4 (1-2236) x Q8WVS2 (1-906)
Qy 7 CCGGTGCGGTCGCGGCTCTTCCAGCTCTTCCAGCTCTTCCAGCGCGGCGGCGGCGGTCG 66
Db 4 ProLeuAspGlyLeuSerValSerSer-----SerThrGlySerLeuGlySer 20
Qy 67 TCACACGACGAGCTGGACCTGGCCCGACCATGGACCGGAAAGTGGCCGCGGAGAAATCCGG 126
Db 21 AlaAlaGlyAlaGlyGlyGlyGlyAlaGlyLeuArgLeuLeuSerAlaAsnValArg 40
Qy 127 -----CATAGAGTGGATTTCTGATGCAATGATGAGAGAGGACTATCTATGAT 180
Db 41 GlnLeuHisGlnAlaLeuThrAlaLeuLeuSerGluAlaGluArgGluGlnPheThrHis 60
Qy 181 GTGCTGCGAATGTACACACGACCATGGACGTCGCGGCTGCTGGGAGACTGAAGCTG 240
Db 61 CysLeuAsnAlaTyrHisAlaArgAsnValPheAspLeuValArgThrLeuArgVal 80
Qy 241 GTCATCAATGAACCCAGCCGCTGCTGCTGTTGATGCCATTCGCGCGCTGATCCCATG 300
Db 81 LeuLeuAspSerProValLysArgLeuLeuProMetLeuArgLeuValIleProArg 100
Qy 301 AAGCACGAGTGGATATGATCAG-----CTGACC 330
Db 101 SerAspGlnLeuLeuPheAspGlnTyrThrAlaGluGlyLeuTyrLeuProAlaThr 120
Qy 331 CCGCGCGCGC-----TCCAGGAAGCTGAAGGAGGTGCGTCTG 366
Db 121 ProTyrArgGlnProAlaTrpGlyGlyProAspSerAlaGlyProGlyGluValArgLeu 140
Qy 367 GACCGTCTGCAC-----CCCGAAGCCCTCGCCCTGAGTGTGCGTGGCGCTG 414
Db 141 ValSerLeuArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySer 160
Qy 415 GAGTTTGGCTGGGCTTTCATCTCCACCTCATCAAGCGGTGAGGAGGAGGAGGAGGAG 474
Db 161 GluHisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGlu 180
Qy 475 GGGCTCCAGGTAGGGGACGAGATCGTCGGGATCAATGGATATTCATCTCCCTCTGATACC 534
Db 181 GlyLeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThr 200
Qy 535 CATGAGGAGGTCAACCTTCTGCAACCAAGAAAGAACTGTGTCCATCAAGTGAAGACAC 594
Db 201 HisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSer 220
```

```
QY 595 ATCCGCTGATCCCGTGAAGAGCTCTCTCTGATGAGCCCTACCTGTCGATGATGATGATGATGATGAT 654
Db 221 AlaGlyArgIleProGlyGlyPyrValThrAsnHisIleTyrThrTriP-----ValAsp 238
QY 655 -----CAGTTTGTGTCGGAATCTGGGCGCTGCGAGGAGCGCTGGGC----- 696
Db 239 ProGlnGlyArgSerIleProSerGlyLeuProGlnProHisGlyGlyAlaLeu 258
QY 697 -----TCCCTTGGAAATCGGGAAC-----AAGGAGAAG 726
Db 259 ArgGlnGlnGlyAspArgArgSerThrLeuHisLeuGlnGlnGlyAspGluLys 278
QY 727 AAGTCTTCATCAGCTGATGAGCTCCGAGGCGCTTGGCTGCGAGCATTTCCAGCGGCC 786
Db 279 LysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAla 298
QY 787 ATCCAGAACCTCGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGCTGCTGAGGTG 846
Db 299 GluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySer 318
QY 847 GGATTGGAGATAGGGACAGATTTGCGAAGTCAATGGCTGCGACTTCTTCACTGGAT 906
Db 319 GlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeu 338
QY 907 CACAGGAGCTGTAATCTGCTGAAATAGCCGAGCTGACCATCTCCATTTGATGCT 966
Db 339 HisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr--ValLys 357
QY 967 GCAGCTGGCGCG-----GAGCTGTTTCATCAGACAGCGGAGCGCTGCGCAGAG 1014
Db 358 AspValGlyArgLeuProHisAlaArgThrThrValAspGluThrIleAlaSer 377
QY 1015 GCGCGCAGCGTGAAGTGGCGGAGGAGCTTCTCATCAGAGCGGCTGCGATGGAG 1074
Db 378 SerArgIleArgGluThrMetAlaAsnSerAlaGlyPheLeuGlyAspLeuThrThrGlu 397
QY 1074 ----- 1074
Db 398 GlyIleAsnLysProGlyPheTyrLysGlyProAlaGlySerGlnValThrLeuSerSer 417
QY 1075 -----TCCAAAGATCTCCAGGAGCAGCAGAGATGGAGCGCAAGAGGAGAAA 1125
Db 418 LeuGlyAsnGlnThrArgValLeuLeuGluGluGlnAlaArgHisLeuLeuAsnGluGln 437
QY 1126 GAAATTTGCCCAAGCAGCAGCAGAGGAAATAGAGATACCG----- 1167
Db 438 GluHisAlaThrMetAlaTyrTyrLeuAspGlu--TyrArgGlyGlySerValSerVal 456
QY 1168 -----AAGAGATGGAAACAGATTTAGAGGAGGAGAGAGAGTAAAGAGCAA 1215
Db 457 GluAlaLeuValMetAlaLeuPheLysLeuLeuAsnThrHisAlaLysPheSerLeuLeu 476
QY 1216 TGGGAA-----GAAGACTGGGCTCAAAGGACAGCTACTCTTG 1254
Db 477 SerGluValArgGlyThrIleSerProGlnAspLeuGluArgPheAspHisLeuValLeu 496
QY 1255 CCTAAACCATCATCTGCTGAGGTACCCAGTACCCCTTCGCAAGCCA----- 1302
Db 497 ArgArg-----GluIleGluSerMetLysAlaArgGlnProGlyProGly 512
QY 1303 -----AAGTAT-----GATCAG 1314
Db 513 AlaGlyAspThrTyrSerMetValSerTyrSerAspThrGlySerSerThrGlySerHis 532
QY 1315 GGATGTGAACCTCAGCTCGAGCCCGCA--GATGACTGTGATGAGCAGGAGCAGCAG 1371
Db 533 GlyThrSerThrThrValSerSerAlaArgAsnThrLeuAsp-----LeuGluGluThr 550
QY 1372 GGAGAG-----CAGGATTTCCCG 1389
Db 551 GlyGluAlaValGlnGlyAsnIleAsnAlaLeuProAspValSerValAspValArg 570
QY 1390 AAATATGAGGAGGCTTTGACCCCTACTCTATGTTCT-----ACCCAGAGCAGATCATG 1443
```

```
Db 571 SerThrSerGlnGlyLeuSerSerPheLysProLeuProArgProProLeuAlaGln 590
QY 1444 GGAAGAGATGTCGGCTCTACGATCAAGAAGAGGAGGATCCTTAGACCTGGCCCTGGAA 1503
Db 591 GlyAsnAspLeuProLeuGlyGlnProArgLysLeuGlyArgGluAspLeuGlnProPro 610
QY 1504 GCGGCTGTGGACTCCCACTGGGAAGGTGCTTCTGCTGCTGATGAGCGGGGAGCT 1563
Db 611 SerSerThrProSerCysSerGly--ThrValPheSerAlaProGlnAsnArgSerPro 629
QY 1564 GCTGAGCGGCATGTGGCATTTGTGAAGGGGAGGAGATCATGGCAATCAACGGCAAGATT 1623
Db 630 Pro-AlaGly----- 632
QY 1624 GTGACAGACTACACCTCGTGTGCTGAGCTGACGCTGCCCTGCAGAGAGGCTGGAATCAGGGC 1683
Db 633 -----ThrAlaPro-----ThrPr 637
QY 1684 GGGAGCTGGATCGACCTTGTGTTGCGCTGTGCCCCCAAGAGGAGTATGACGATGAGCTG 1743
Db 637 oGlyThrSerSerAlaGlnAspLeuProSerSer--Pro----- 649
QY 1744 ACCTCTGCTGAACTCCAAAGGGGAAACCAATTCACGGCTTAGGAAACAGTCAGCTC 1803
Db 650 --IleTyrAlaSerValSerProAlaAsnProSerSerLys----- 662
QY 1804 CGGCCCCACCTCGTGACACACAAAGCCTCGGACCAGCCTTGAGAGAGGCCACATCAGACAC 1863
Db 663 -----ArgProLeuAspAlaH 668
QY 1864 ACCAGATGGCATCTTGGGAGCTGAATCTATCATCCAGGAGTCTCAAACTCCCTTTGGCC 1923
Db 668 1s-----LeuAlaLeuValAsnGlnHisProIleGlyP 679
QY 1924 CTGAACCA 1931
Db 679 roPhePro 681
RESULT 9
Q9H5P4 PRELIMINARY; PRT; 517 AA.
AC Q9H5P4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CDNA: FLJ23209 fis, clone ADH00512 (Hypothetical 55.7 kDa protein).
DE Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADIPOSE TISSUE;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026862; BAB15577.1; -
DR EMBL; BC015692; AALH5692.1; -
DR HSSP; Q12923; 3PDZ.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
KW Hypothetical protein.
```

SQ SEQUENCE 517 AA: 55676 MW; 06C9C614283478F7 CRC64;

Alignment Scores:

Pred. No.: 3,52e-11 Length: 517
Score: 283.50 Matches: 104
Percent Similarity: 42.71% Conservative: 63
Best Local Similarity: 26.60% Mismatches: 162
Query Match: 6.88% Indels: 62
DB: 4 Gaps: 11

US-09-502-945-4 (1-2236) x Q9H5P4 (1-517)

Qy 349 CTGAAGAGGTCGTCGTGGACCGTCGCACCCGAGGCTCGGCTGAGTGTGGTGT 408
Db 84 lIeHISerValArgValGluLysSerProAlaGlyArgLeuGlyPheSerValArgGly 103
Qy 409 GGCCTGGAGTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAAAGGCGTCAGCAGAC 468
Db 104 GlySerGluHisGlyLeuGlyIlePheValSerLysValGluGluGlySerSerAlaGlu 123
Qy 469 AGCGTCGGCTCCAGGTAGGAGCAGATCGTCGGATCAATGGATATTCATCTCTCC 528
Db 124 ArgAlaGlyLeuGlyValGlyLysPheValSerLysValGluGluGlySerSerAlaGlu 143
Qy 529 TGTACCCATGAGGAGTCTCATCAACCTCATTCGAACCAAGAAACTGTGTCCATCAAGTG 588
Db 144 ThrThrMetGlySerAlaValLysValLeuThrSerSerArgLeuHisMetMetVal 163
Qy 589 AGACATCGGCTGATCCCGGTGAAAGCTCTCTGATGAGCCCTCATCTGTGGCAGTAT 648
Db 164 ArgArgMetGlyArgValProGlyIleLysPheSerLysGluLysThrTrp----- 181
Qy 649 GTGGAT-----CAGTTTGTGCGAATCTGGGGCTGGAGGAGCGCTGGGC 696
Db 182 ValAspValAlaSerArgLeuValGluLys-----CysGlySerThrPro 198
Qy 697 TCCCTGGAAATCGGGAACCAAGGAGAGGCTTCATCAGCGCTGGTGGCTCCGGA 756
Db 199 SerAspThrSerSerGluAspGlyValArgGlyIleValHisLeuThrThrSerAsp 218
Qy 757 GGC-----CTTGGCTGAGCATTTCCAGCGCCCTCCAGAACCGTGGCATCTTTATC 810
Db 219 AspPheCysLeuGlyPheAsnIleArgGlyLysGluPheGlyLeuGlyIleThrVal 238
Qy 811 AGCCATGTGAACCTGGCTCCTCTGCTGAGTGGGATTCGAGATAGGAGCAGCATTT 870
Db 239 SerLysValAspHisGlyLeuAlaGluGluAsnGlyIleLysValGlyAspGlnVal 258
Qy 871 GTCGAAGTCAATGGCTCGACTTCTTAACCTGGATCAAGGAGGCTGTAATGTGCTG 930
Db 259 LeuAlaAlaAsnGlyValArgPheAspPheSerHisSerGlnAlaValGluValLeu 278
Qy 931 AAAATAGCCGAGCTGACCATCTCCATTTAGTCTGAGTGGGATTCGAGATAGGAGCAGCAT 990
Db 279 LysGlyGlnThrHisIleMetLeuThrIle---LysGluThrGlyArg-----TyrPro 295
Qy 991 ACAGACGGGAGCGCTCGAGCGCGGAGCGCTGAGCTGACCGCGCAGGAGCTTCTC 1050
Db 296 AlaTyrLysGluMetValSerGluTyrCysTrpLeuAspArgLeuSerAsnGlyValLeu 315
Qy 1051 ATGCAGACGGCTGGCGATGGATGCCAACAGATCTCCAGGAGCAGCAGGAGATGGAG 1110
Db 316 GlnGlnLeuSerProAlaSerGluSerSerSerSerValSerSerCysAlaSerSerAl 335
Qy 1111 CGGCAAGAGAGAAATATGCCAGAGGCGAG-----CAGAGGAA 1152
Db 335 aProTyrSerSerGlySerLeuProSerAspArgMetAspPheCysLeuGlyGlnGlu 355
Qy 1153 AATGAGATACCGGAGGAGATGGAACAGA----- 1183
Db 355 uProGlySerArgGlyProGlyTrpGlyArgAlaAspThrAlaMetGlnThrGluProAs 375
Qy 1184 -----TTGTAGAGGAGGAGAG 1200

Db 375 pAlaGlyGlyArgValGluThrTrpCysSerValArgProThrValIleLeuArgAspTh 395
Qy 1201 AAGTTTAAGAACCAATGGGAAGAGACTGGGCTCAAGGAGCAGCTACTCTTGCCTAAA 1260
Db 395 rAlaIleArgSerAspGlyProHisProGlyArgArgLeuAspSerAlaLeuSerGluSe 415
Qy 1261 ACCA-----TCACTGTGTGAGGTACACCCAGCTACCC-----TTCCGAAGCCA 1302
Db 415 rProLysThrAlaLeuLeuAlaLeuSerArgProArgProLleThrArgSerGl 435
Qy 1303 AAGTATGATCAGGAGGTGGAACCTGAGCTGAGCCGAGCAGCTGAGTGGATGGAGGACG 1362
Db 435 nSerTyrIleThrLeuTrp-----GluGluLy 444
Qy 1363 GAGGAGCAGGAGGAGCAGGAGTTCCTCCGAAAT 1393
Db 444 sGlnGlnArgLysLysGluLysSerGlySer 454
RESULT 10
Q921G9 PRELIMINARY; PRT; 904 AA.
ID Q921G9; AC Q921G9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to tight junction protein 3.
GN TJP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RX NCBI_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012518; AAI12518.1; -
DR MGD; MGI:1351650; TJP3.
DR InterPro; IPR000619; Guanylate_kin.
DR Pfam; PF00625; Guanylate_kin; 1.
DR PRINTS; PF00595; PDZ; 3.
DR PRINTS; PRO1597; ZONOCCLUDNS.
DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 904 AA; 99169 MW; A4D9C58AEEC485CC CRC64;
Alignment Scores:
Pred. No.: 1.7e-10 Length: 904
Score: 275.00 Matches: 134
Percent Similarity: 34.51% Conservative: 90
Best Local Similarity: 20.65% Mismatches: 191
Query Match: 6.67% Indels: 234
DB: 11 Gaps: 24
US-09-502-945-4 (1-2236) x Q921G9 (1-904)
Qy 385 GGCCTCGCCTGAGTGTGCTGGCTGGAGTTGGCTGTGGG---CTCTTCATCTCC 441
Db 21 GlyPheGlyIleAlaValSerGlyGlyHisAspArgAlaSerGlySerValValSer 40
Qy 442 CACCTCATCAAGGCGGTTCAGCAGCAGCAGCGGTCCGGCTCCAGGTAGGGAGAGATCGTC 501
Db 41 AspValValProGlyGlyProAlaGluGly---ArgLeuArgThrGlyAspHisIleVal 59
Qy 502 CGGATCAATGGATATTCATCTCTCTGATCCCATGAGGAGTCAACCATTCATTCGA 561
Db 60 MetValAsnGlyValSerValGluAsnValThrSerAlaPheAlaIleGlnIleLeuLys 79
Qy 562 ACC---AAGAAACTGTGTCTCATCAAAAGTG-----AGACATCGCGCTGATCCCC 609
Db 80 ThrCysThrLysThrAlaAsnValThrValLysArgProArgValGlnLeuProAla 99

QY 610 GTGAAAGCTCTCTGATGAGCCCTCACTTGGCAGTATGTGGATCACTTGTGTGCGAA 669
 Db 100 ThrLysAlaSerProAlaSer-----GlyHisGlnLeuSerAspGlnGluAlaAsp 117
 QY 670 TCTGGG----- 675
 Db 118 HisGlyArgGlyTyrGluGlyAspSerSerGlySerGlyArgSerTrpGlyGluArg 137
 QY 676 -----GGCTGGCGAGCGACCTGGCTCCCTCGGAAATCGGAAAC 717
 Db 138 SerArgArgSerArgAlaGlyArgArgGlyArgValGlySerHisGlyArgArgSerSer 157
 QY 718 -----AAG 720
 Db 158 GlyGlyGlySerGluAlaAsnGlyLeuAspLeuValSerGlyTyrLysArgLeuProLys 177
 QY 721 GAG-----AAGAGGTCTTCATCAGCCTGGTAGCTCCGAGGC 759
 Db 178 GlnAspValLeuMetArgProLeuLysSerValLeuValLysArgArgAsnSerGluGlu 197
 QY 760 CTTGGCTGCAGCATTTCCAGCGCCCTCCAGAGCCTGGCATCTTTATCAGCCATGTG 819
 Db 198 PheGlyValLysLeuGlySer-----GlnIlePheIleLysHisIle 211
 QY 820 AAACCTGGCTCCTGTCTGCT---GAGGTGGATTGGATAGGAGCAGCATTTGCGAA 876
 Db 212 ThrGluSerGlyLeuAlaAlaArgAsnHisGlyLeuGlnGluGlyAspLeuIleLeuGln 231
 QY 877 GTCAATGGCTGCATCTCTTAACCTGGATCAGAGAGGCTGTAAATGTGCTGAAAT 936
 Db 232 IleAsnGlyValSerSerAlaAsnLeuSerLeuSerAspThrArgArgLeuIleGluLys 251
 QY 937 AGCCGC---AGCTGACCATCTCCATTGTAGCTGCAGCTGGCGGAGCTG----- 984
 Db 252 SerGluGlyGluLeuThrLeuLeuValLeuArgAspSerGlyGlnPheLeuValAsnIle 271
 QY 985 -----TTCATCACAGAC-----CGGGAG 1002
 Db 272 ProProAlaValSerAspSerAspSerLeuMetGluAspIleSerAspLeuThrSer 291
 QY 1003 CGGTGGCAGAGCGCGGCGGAGCTGTAGCTGCAGCGCAGGAGCTCTCATGCAAGCGG 1062
 Db 292 GluLeuSerGlnAlaProProSerHisValProProProLeuLysGlyGln---Arg 310
 QY 1063 CTGGCGATGAGTCCAAAGATCCTCCAGGACGACGAGATGGAGCGCAAGAGGAGA 1122
 Db 311 SerProGluAspSerGlnThrAspSerProValGluThrProGlnProArgArgArgGlu 330
 QY 1123 AAGAAATTTCCAGAGGAGCAGAGGAAATGAGAGATACCGGAAGGAGATGGAACAG 1182
 Db 331 ArgSerValAsnSerArgAlaIleAlaGluProGluSer----- 343
 QY 1183 ATTGTAGAGGAGGAGAGAGATTAAAGACAAATGGAAGAGACTGGGGCTCAAAGGAA 1242
 Db 343 ----- 343
 QY 1243 CAGCTACTCTTGCTAAACCATCACTGCTGAGGTACACCCAGTACCTTCGCAAGCCA 1302
 Db 344 -----ProGlyGluSerArgTyrAspIleTyrArgValProSerArgGlnSer 359
 QY 1303 AAGTATGATCAGGAGTGGAACTGAGCTGAGCCGCGCAGATCACCTGGATGGAGCAGG 1362
 Db 360 LeuGluAspArgGlyTyrSerPro----- 367
 QY 1363 GAGGAGCGGAGAGCAGGAGTTCCGGAANAATAGAGAAAGCTTTGACCCCTACTCTATG 1422
 Db 367 ----- 367
 QY 1423 TTCACCCACAGCAGATCATGGGAAGGATGTCGGCTCCTACGCATCAGAGAGGAGGA 1482
 Db 368 -----AspThrArgValValSerPheProLysGlyAla 378

QY 1483 TCCTTAGACCTGGCCCTGGGAAGCGGTGTGAGACTCCCCCATTTGGGAAGGTGGTCTTCT 1542
 Db 379 SerIleGlyLeuArgLeuAlaGlyGlyAsnAsp-----ValGly---IlePheValSer 395
 QY 1543 GCTGTGTATGAGCGGCGGAGCTGCTGAGCGCATGTGGCATTTGTGAAGGGAGCAGCATC 1602
 Db 396 GlyVal---GlnAlaGlySerProAlaAspGlyGlnGlyIleGlnGluGlyAspGluIle 414
 QY 1603 ATGGCAATCAACGCGCAAGATTGTGACAGACTACACCTGGCTGAGGCTGAGCTGCCCTG 1662
 Db 415 LeuGlnValAsnGlyMetProPheArgAsnLeuThrArgGluGluAla----- 430
 QY 1663 CAGAGGCTGGAATCAGGCGGCGGACTGGATCCACCTTGTGGTGTGGCTGTCGCCCCA 1722
 Db 431 -----ValGlnPheLeuLeuGlyLeuProProGly 440
 QY 1723 AAGGAGTATGAC-----GATGAGCTGACCTTCTTCTGTAAGTCC 1761
 Db 441 GluAspMetGluLeuValThrGlnSerLysGlnAspIlePheLysLysMetValGlnSer 460
 QY 1762 AAAAGGGGAAACCAATTCACGCGCTTAGGAACACAGTACGAGTCCGCGCC----- 1809
 Db 461 ArgValGlyAspSerPheTyrIleArgThrHisPheGluLeuGluProSerProProTyr 480
 QY 1810 -----CACCTCGTAACACA---AACCTCGGACC 1836
 Db 481 GlyLeuGlyPheThrArgGlyAspValPheHisValValAspThrLeuTyrProGlySer 500
 QY 1837 AGCTTGAGAGCGCCACATCACACACAGATGGCATCTTGGACCTGAATCTATCA 1896
 Db 501 GlyProGlyHisGlyHisSerHis----- 509
 QY 1897 CCCAGAACTCAAACTCCCTTGGCCCTGAACACGAGCCAGATGAAGAACAGCTCGGC 1956
 Db 509 ----- 509
 QY 1957 CACTTTTGAAGCCAAATGTGGAGAAAGGAGAGCAGCCGTTGGGAGAGATCTC 2016
 Db 510 -----GlyGlyLeuTrpLeuAlaAlaArgMetGlyArgAspLeu 522
 QY 2017 AAGGATCCAGACTCTCATTTCTTCTTCT 2043
 Db 523 ArgGluGlnGluArgGlyValIlePro 531
 RESULT 11
 Q9NTP3 PRELIMINARY; PRT; 202 AA.
 AC Q9NTP3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE BA108L7.3 (Novel PDZ (DHR, GLGF) domain protein) (Fragment).
 GN BA108L7.3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Heath P.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL13215; CAB92801.1; -
 DR HSSP; P29476; 10AV.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 2.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 FT NON_TER 1
 FT NON_TER 202
 SQ SEQUENCE 202 AA; 21709 MW; 50E6507836F8845C CRC64;
 Alignment Scores:
 Pred. No.: 1.51e-10 Length: 202

Score: 272.00 Matches: 69
Percent Similarity: 52.74% Conservative: 37
Best Local Similarity: 34.33% Mismatches: 83
Query Match: 6.60% Indels: 12
DB: 4 Gaps: 5

US-09-502-945-4 (1-2236) x O9NTP3 (1-202)

QY 379 CCCGAAGGC--CTCGGCTGAGTGGTGGTGGCCCTGGAGTTGGTGGGCTCTTC 435
Db 6 ProAlaGlyArgLeuGlyPheSerValArgGlyGlySerGluHisGlyLeuGlyPhe 25
QY 436 ATCTCCACCTCATCAAGAGCGGTGAGGACAGCGTCCGGCTCCAGGTAGGGACGAG 495
Db 26 ValSerLysValGluGlySerSerAlaGluArgAlaGlyLeuGlyValGlyAspLys 45
QY 496 ATCTCCGCGATCAATGGATATTCATCTCCCTGCTATCCCATGAGGAGGTCAATCACTC 555
Db 46 IleThrGluValAsnGlyLeuSerLeuGluSerThrThrMetGlySerAlaValLysVal 65
QY 556 ATTCGAACCAAGAAACACTGTCTCCATCAAGTGAGACACATCGGCTGATCCCGTGAAA 615
Db 66 LeuThrSerSerArgLeuHisMetMetValArgArgMetGlyArgValProGlyIle 85
QY 616 AGCTCTCTGTATGAGCCCTCACTTGGCAGTATGTGAT-----CAGTTTGTG 663
Db 86 LysPheSerLysGluLysThrTrp-----ValAspValValAsnArgLeuVal 103
QY 664 TCGGAATCTGGGGCGTGGAGGAGCTGGCTCCCTCGAAATCGGGAAACAGGAG 723
Db 104 ValGluLys-----CysGlySerThrProSerAspThrSerSerGluAspGlyVal 120
QY 724 AAGAGGTCTTCATCAGCCTGTAGGCTCCGAGGC-----CTGGCTGAGCATTTCC 777
Db 121 ArgArgIleValHisLeuThrThrThrSerAspPheCysLeuGlyPheAsnIleArg 140
QY 778 AGCGGCCCATCCAGAACCTGGCATCTTTATACACCATGTGAACCTGGCTCCCTGCT 837
Db 141 GlyGlyLysGluPheGlyLeuGlyIleTyrValSerLysValAspHisGlyGlyLeuAla 160
QY 838 GCTGAGTGGGATGGAGATGGGACAGATTTGCGAAGTCAATGGCGTCACTTCTCT 897
Db 161 GluGluAsnGlyIleLysValGlyAspGlnValLeuAlaAlaAsnGlyValArgPheAsp 180
QY 898 AACCTGGATCAAGAGGCTGTAATGTCTGCTGAAATATGCCGAGCTGACCATCTCC 957
Db 181 AspIleSerHisSerGlnAlaValGluValLeuLysGlyGlnThrHisIleMetLeuThr 200
QY 958 ATT 960
Db 201 Ile 201

RESULT 12
O75970 PRELIMINARY; PRT; 2042 AA.

ID O75970
AC O75970;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Multi PDZ domain protein MUPP1.
GN MUPP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
FP
FC
PP
PT
RA Eng L., Kravinsky G., Clapham D. E.;
RT "Human homolog of MUPP1 protein."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF093419; AAC61870.1; -.
DR HSSP; Q12959; 1PDR.

DR InterPro; IPR004172; L27.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF02828; L27; 1.
DR Pfam; PF00595; PDZ; 13.
DR SMART; SM00228; PDZ; 13.
DR PROSITE; PS00106; PDZ; 13.
SQ SEQUENCE 2042 AA; 218595 MW; F0E1DD029749A762 CRC64;

Alignment Scores:
Pred. NO. 5.65e-10 Length: 2042
Score: 269.50 Matches: 145
Percent Similarity: 36.23% Conservative: 76
Best Local Similarity: 23.77% Mismatches: 170
Query Match: 6.54% Indels: 219
DB: 4 Gaps: 26

US-09-502-945-4 (1-2236) x O75970 (1-2042)

QY 11 TCGGGTGGCGGCTTTCACAGCTCTGGCAGCGGGCCAGAA-----GGAACG 61
Db 1585 SerGlnSerLeuMetValProGlnSerGlySerProGluProGluSerIleArgAsnThr 1604
QY 62 GGTCTGCAACGACGACGCTGGACCTGGCCAGCCATGGACGAAAGTGGCCGAGAAT 121
Db 1605 SerArgSerThrProAlaIlePheAla-Ser---AspProAlaThrCysProIleI 1623
QY 122 TCCGGCATNAGGTGGATTTCTGATTGAAATGATGC----- 158
Db 1623 eProGly-----CysGluThrThrIleGluIleSerLys 1634
QY 159 ---AGAGAAGGACTATCTCTATGATGTGTG----- 186
Db 1634 sGlyArgThrGlyLeuGlyLeu-SerIleValGlyGlySerAspThrLeuGlyAlaI 1654
QY 187 -----CGAATGTACCACAG-----ACCATGGAGCTGGCGTCTCGTGG 226
Db 1654 leIleIleHisGluValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaG 1674
QY 227 GAGACCTGAGCTGGTCACTCAAT----- 249
Db 1674 LysAspGlnIleLeuGluValAsnGlyIleAspLeuArgLysAlaThrHisAspGluAlaI 1694
QY 250 -----GAACCCAGCGCTGCTCCCTCTGTTGTGATGCCATTCGGCCGCTGA 292
Db 1694 leAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyrArgAspGluA 1713
QY 293 TCCCATGAAGCACCAGGTGGAAATATGATCAGCTGACCCCGCGCGCTCCAGGAAGCTGA 352
Db 1713 laProTyrLysGluGluValCysAspThrLeuThr----- 1725
QY 353 AGGAGGTGGCTGGACCGCTGTCACCCCGAGGCGCTGGCGCTGAGTGTGGTGGGCC 412
Db 1726 -----IleGluLeuGlnLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA 1744
QY 413 TGGAGTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGGCGCTCAGCAGACAGCG 472
Db 1744 rgAsn---AspThrGlyValPheValSerAspIleValLysGlyIleAlaAspAlaA 1763
QY 473 TCGGG---CTCCAGGTAGGGACGAGATCGTCGGGATCAATGGATATTCATCTCTCTCT 529
Db 1763 spGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGlyGluAspValArgAsnA 1783
QY 530 GTACCCATGAGGAGGTCACTCAACTCATTCGA---ACCAAGAAACCTGTGTCCATCAAG 586
Db 1783 laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGlyThrValThrLeuGluV 1803
QY 587 TGACACATCGGCTGATCCCCGTGAAAGCTCTCTCTGATGAGCCCTCCTACTTGGCAGT 646
Db 1803 al-----GlyArgIleLysAlaGlySerThrSerGluSerLeu----- 1816
QY 647 ATGTGGATCAGTTTGTGCGGAATCTGGGGCGTGGCAGGCGCTGGGTCC-----C 700
Db 1817 -----GluSerSerLysLysAsnAlaLeuAlaSerGluIleG 1830

```
QY 701 CTGGAAATCGGAAACAAAGGAGAGAGTCTTATCAGCGCTGGTAGCTCCCGAGGCC 760
Db 1830 InGlyLeuArgThrValGluMetLysLys-----GlyProThrAspSerL 1845
QY 761 TTGGCTGCAGCATTTCCAGCGC-----CCCATCCAGAAGCCTGGCATCTTTATCA 811
Db 1845 euGlySerIleAlaGlyValGlySerProLeuGlyAspValProIlePheIleA 1865
QY 812 GCATGTGAAACCTGGCTCCCTGCTGCTGAGGTG---GGATTGGAGATAGGGACACGA 868
Db 1865 IamMetHisProThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgI 1885
QY 869 TTGTCGAAGTCAATGGCTCGGACTTCTTAACCTGGATCACAAGAGAGCTGTAAATGTC 928
Db 1885 IeValThrIleCysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValAsnLeuL 1905
QY 929 TGAATAATAGCGGACGCTGACCATCTCCATGTAGCTGCAGCTGGCGGGAGCTGTTC 988
Db 1905 euLysAsnAla---SerGlySerIleGluMetGlnValAlaGlyGlyAspVal--- 1922
QY 989 TCAGACACGGGAGCGGCTGGCGAGAGCGCGCAGCGTGCAGCTGCAGCGGAGGAGTTC 1048
Db 1923 -----SerValValThrGlyHisGlnGlnGluProAlaSerSerLeus 1938
QY 1049 TCATCAGAACGGCTGGCGATGGAGTCCACAAGATCTCCAGGAGCAGGAGATGG 1108
Db 1938 exPhe-----ThrGlyLeuThrSerSerIlePheGln----- 1949
QY 1109 AGCGGCAAGAGAGAAAGAAATGCCAGAACGACGAGAGAAATGAGAGATACCGGA 1168
Db 1949 ----- 1949
QY 1169 AGGAGATGAACAGATTTGTAGAGAGAGAGAGATTTAAGAAGCAATGGGAAGAGACT 1228
Db 1950 -----AspAspL 1952
QY 1229 GGGGCTCAAGGAACAGTACTTCTTGCCTAAACATCATCTGCTGAGGTACACCCAGTAC 1288
Db 1952 euGlyProGln-----CysLysSerIleThrLeuGluArgGlyPro--- 1966
QY 1289 CCCTTCGCAAGCAAAGTATGATCAGGGAGTGAACCTCAGCTCGAGCCCGCAGATGACC 1348
Db 1966 ----- 1966
QY 1349 TGGATGGAGCAGGAGGAGCAGGAGAGAGATTCGGAATATGAGGAAGGCTTG 1408
Db 1967 --AspGly----- 1968
QY 1409 ACCCTACTCTATGTTACCCCGCAGCAGATCATGGGGAGGATGTCGGGCTCCTACGCA 1468
Db 1968 ----- 1968
QY 1469 TCAAGAAGGAGGATCCTTAGACTGGCCCTGGAAGCGGTGTGGACTCCCGCATTTGG- 1527
Db 1969 -----LeuGlyPheSerIleValGlyGlyThrGlySerProHisGlyA 1983
QY 1528 -----AAGTGTGCTGTTCTGCTGATCAGCGGGAGCTGCTGACGGGCATGTGGCA 1582
Db 1983 spLeuProIleTyValLysThrValPheAlaLysGlyAlaAlaSerGluAspGlyArgL 2003
QY 1583 TTGTGAAGGGACGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTGG 1642
Db 2003 euLysArgGlyAspGlnIleIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisG 2023
QY 1643 CTGAGCTGAGCTGCGCTGCAGAG 1668
Db 2023 luGluAlaValAlaIleLeuLysArg 2031
RESULT 13
ID O55164 PRELIMINARY; PRT; 2054 AA.
AC O55164;
```

```
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Multi PDZ domain protein 1.
GN MUPP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98136865; PubMed=9537516;
RA Ullmer C., Schmuck K., Figge A., Lubbert H.;
RT "Cloning and characterization of MUPP1, a novel PDZ domain protein.";
RL FEBS Lett. 424:63-68(1998).
EMBL: AJ001320; CAA04681.1; .
DR HSP; Q12959; IPRD.
DR InterPro; IPR004172; I27.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF02828; L27; 1.
DR Pfam; PF00595; PDZ; 13.
DR SMART; SM00228; PDZ; 13.
DR PROSITE; PS01016; PDZ; 13.
SQ SEQUENCE 2054 AA; 218591 MW; 44BD3F42B801F78F CRC64;

Alignment Scores:
Pred. No.: 6,13e-10 Length: 2054
Score: 269.00 Matches: 120
Percent Similarity: 36.64% Conservative: 72
Best Local Similarity: 22.90% Mismatches: 160
Query Match: 6.53% Indels: 172
DB: 17

US-09-502-945-4 (1-2236) x O55164 (1-2054)
QY 157 GCAGAGAGGACTATCTATGATGCTGCGAATGTACACACAGACCATGAGCTGGGCC 216
Db 1672 AlaThrHisAspGluAlaIleAsnValLeuArgLThrProGlnArgVal----- 1688
QY 217 GTGTCGTGGGAGAGCCTGAAGCTGGTATCAATGAACCCAGCCGCTCTGCTCTGTTGAT 276
Db 1689 -----ArgLeuThrLeuTyr--- 1693
QY 277 GCCATTCGGCGCTGATCCACTGAAGCACAGGTGGATATGATCAGCTACCCCGCG 336
Db 1694 -----ArgAspGluAlaProTyLysGluAspValCysAspThrPheThr----- 1709
QY 337 CGCTCCAGGAAGCTGAAGGAGGTGCTGACCGTCTGCACCCGCTGCGAGGCTCGGCTG 396
Db 1710 -----ValGluLeuGlnLysArgProGlyLysGlyLeuGlyLeu 1722
QY 397 AGTGTGCTGGTGGCTGGAGTTTGGCTGTGGCTCTTCATCTCCCATCATCAAGGC 456
Db 1723 SerIleValGlyLysArgAsn---AspThrGlyValPheValSerAspIleValLysGly 1741
QY 457 GGTCCAGGACAGACCGCTCGG---CTCCAGGTAGGGAGAGATCGTCCGATCATGGA 513
Db 1742 GlyIleAlaAspAlaAspGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGly 1761
QY 514 TATTCATCTCTCTCTGATCCCATGAGGAGGTTCATCAACCTCATTCGA---ACCAAGAA 570
Db 1762 GluAspValArgAsnAlaThrGlnGluAlaValAlaLeuLeuLysCysSerLeuGly 1781
QY 571 ACTGTGTCATCAAGAGTACACATCGGCTGATCCCGCTGATCCCGTGAAGAGCTCTCTGATGAG 630
Db 1782 ThrValThrLeuGluValGlyArgIleLysAlaAlaProPheHisSerGluArgArgPro 1801
QY 631 CCCCTCACTTGGCAGTATGTGATGAGTTCGCGGATCTGCGGAGTCTGGGGCGCTCGGAGCAGC 690
Db 1802 SerGlnSerSerGlnValSerGlnSerLeuSerPheSerLeuProArgSerGly 1821
QY 691 CTGGGCTCCCTGGAAATCGGGAACCAAGAGAGAGAG----- 729
```



```
Db 1822 IieHisThrSerGluSerSerGluSerSerAlaLysLysAsnAlaLeuAlaSerGluLile 1841
Qy 730 -----GTCCTTCATCAGCCCTGGTAGGCTCCCGAGGCGCTTGGCTGCGACGATT 774
Db 1842 GlnGlyLeuArgThrValGluLileLysLysGlyProAlaAspAlaLeuGlyLeuSerLile 1861
Qy 775 TCCAGCGGC-----CCATCCAGAGAGCTGGCACTTTATCAGCCATGTGAACCT 825
Db 1862 AlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetMetHisPro 1881
Qy 826 GCCTCCCTGCTGCTGAGGTG---GATTGGAGATAGGGGACGAGCTTCTCAAGTCAAT 882
Db 1882 AsnGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrLecys 1901
Qy 883 GCGCTGACTTCTTAACCTGGATCACAAGAGGCTGTAAATGTCTGAAAAATAGCCGC 942
Db 1902 GlyThrSerAspGlyMetThrHisThrGlnAlaValAsnLeuMetLysAsnAla--- 1920
Qy 943 AGCCTGACCATCTCCATGCTAGCTGACGCTGCGCGGAGCTGTTATGACACACCGGAG 1002
Db 1921 SerGlySerIleGluValGlnValAlaGlyGlyAspVal----- 1934
Qy 1003 CGCTGCGCAGAGCGCGGCGGAGCTGAGCTGCGAGCGGAGGCTTCTCATGCAAGCGG 1062
Db 1935 SerValValThrGlyHisGlnGlnGluLeuAlaAsnProCysLeuAlaPhe----- 1951
Qy 1063 CTGGCGGAGGAGTCCCAACAGATCTCCAGGAGCAGCAGGAGATGGCGGCAAGGAGA 1122
Db 1952 ThrGlyLeuThrSerSerThrIle----- 1959
Qy 1123 AAGAAATGCCCCAGGACGACGAGGAGGAAATGAGAGTATACCGAAGAGATGGAACAG 1182
Db 1959 ----- 1959
Qy 1183 ATTGTAGAGGAGGAGAGAGAGTTTAAGAGCAATGGAAGAGACTGGGCTCAAGGAA 1242
Db 1960 -----PheProAspLeuGlyProProGln 1968
Qy 1243 CAGCTACTCTGCTTAAACCATCTACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCA 1302
Db 1969 Ser-----LysThrIleThr----- 1973
Qy 1303 AAGTATGATCAGGAGTGAACCTGAGCTGAGCGCGGAGGATGACCTGATGGAGGCGCAG 1362
Db 1974 -----LeuAspArgGlyPro 1978
Qy 1363 GAGGAGCAGGAGAGCAGGATTTCGGAATATGAGGAAGGCTTTGACCCCTACTCTATG 1422
Db 1979 Asp----- 1979
Qy 1423 TTCACCCAGACAGCAGATCATGGGGAAGGATGTCGGCTCTTACGATCAAGAAGAGGGA 1482
Db 1979 ----- 1979
Qy 1483 TCCTTAGACCTGGCCCTGGAAGCGGTGGGACTCCCCCATTTGG-----AAGGTGTC 1536
Db 1980 GlyLeuGlyPheSerIleValGlyGlyTyrglySerProHisGlyAspLeuProIleTy 1999
Qy 1537 GTTCTGCTGCTATGAGCGGGAGCTGCTGAGCGGCATGTTGGCTTCTGTAAGAGGAGC 1596
Db 2000 ValLysThrValPheAlaLysGlyAlaAlaGluAspGlyArgLysArgGlyAsp 2019
Qy 1597 GAGTATGGCAATCAACGGCAAGATTGTGACAGACTACACCCCTGGCTGAGCTGACGCT 1656
Db 2020 GlnIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaValAla 2039
Qy 1657 GCCTGCGAGAAG 1668
Db 2040 IleLeuLysArg 2043
RESULT 14
Q96C69
```

```
ID Q96C69 PRELIMINARY; PRT; 835 AA.
AC Q96C69;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 87.7 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014632; AAH14632.1;
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR PROSITE; PS50106; PDZ; 3.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 835 AA; 87669 MW; F59DF419D7AAF991 CRC64;

Alignment Scores:
Pred. No.: 8,76e-10 Length: 835
Score: 264.50 Matches: 190
Percent Similarity: 32.75% Conservative: 93
Best Local Similarity: 21.99% Mismatches: 233
Query Match: 6.42% Indels: 350
DB: Gaps: 36

US-09-502-945-4 (1-2336) x Q96C69 (1-835)
Qy 289 CTGATCCCACTGAAGCAGCAGGTGGTAAATATGATCAGTCCGCCCGCGCTCCAGGAAG 348
Db 30 IleThrProLeuArgProGluAspAspTyrSerPro-----ArgGluArg 45
Qy 349 CTGAAGGAGGTGCTGCTGGAGCCGCTGCGACCCCGAA----- 384
Db 46 GlyGlyGlyLeuArgLeuProLeuProGluSerProGlyProLeuArgGlnArg 65
Qy 385 -----GGCTCGGCTGAGTGTGCGTGGTGC----- 411
Db 66 HisValAlaCysLeuAlaArgSerGluArgGlyLeuGlyPheSerIleAlaGlyLys 85
Qy 412 -----CTGGAGTTTGGTGTGGCTTCTTCATCTCCACCTCATCAAGGC 456
Db 86 GlySerThrProTyrArgAlaGlyAspAlaGlyIlePheValSerArgIleAlaGlyLys 105
Qy 457 GGTGAGGAGCAGCGTCCGG-----CTCAGTAGGGAGAGATGCTCCGGATCAATGGA 513
Db 106 GlyAlaAlaHisArgAlaGlyThrLeuGlnValGlyAspArgValLeuSerIleAsnGly 125
Qy 514 TATTCATCTCTCTGCTGACCATGAGGAGTCACTCACTCATT---CGAACCAAGAA 570
Db 126 ValAspValThrGluAlaArgHisAlaValSerLeuLeuThrAlaAlaSerPro 145
Qy 571 ACTGTGTCATCAAAAGTGAGACACATCGCTGATCCCGTGAAAGCTCTCCT----- 624
Db 146 ThrIleAlaLeuLeuLeuGluArgGluAlaGlyGlyProLeuProProSerProLeuPro 165
Qy 625 -----GATGAGCCCTCACTTGGCAGTATGTCAGTATGTCGGAATCTGGGGGC 678
Db 166 HisSerProProThrAlaAlaValAlaThrSerIleThrAlaThrProGly 185
Qy 679 GTGCGAGC----- 687
Db 186 ValProGlyLeuProSerLeuAlaProSerLeuAlaAlaLeuGluGlyProTyr 205
Qy 688 -----AGCCTGGCTCCCTCGGAATCGGAAACCAAGAG 723
Db 206 ProValGluGluIleArgLeuProArgAlaGlyGlyPro----- 218
```



```
QY 1410 C-----CCCTACTCTATGTTACCCAGAGCAGATC-----ATGGGGAA 1448
Db 1296 uLeuGluValArgValProGlnAlaGluGlyProProlsArgValSerLeuValGlyAl 1316
QY 1449 GGAT-----GTCCGGCTCTACCATCAAGAGGAGG 1481
Db 1316 aAspAspLeuArgLysMetGlnGluGluAlaArgLysLeuGlnLysArgAlaG1 1336
QY 1482 ATCCCTTAGACCTGGCCCTGGAAGGGGGTGTGGACTCCCCCATTTGGGAAGGTGTCGTTTC 1541
Db 1336 nMetLeuArgGluAlaAlaGluAlaGlyAlaGluAlaArg----- 1349
QY 1542 TGCTGTGTATGACGGGGGAGCTGCTGAGCGCATGTGGCATTTGTCAAAGGGGACGAGAT 1601
Db 1349 ----- 1349
QY 1602 CATGGCAATCAAGGCAAGATTGTGACAGACTACACCTGGCTGAGGTGACGCTGCCCT 1661
Db 1350 -LeuAlaLeuAspGlyGlu-----ThrLeuGlyGlu----- 1359
QY 1662 GCAGAAGGCTGGAATCAGGGGGGAGCTGGATCGACCTTGTGTTGCGTCTGCCCCOC 1721
Db 1359 ----- 1359
QY 1722 AAAGGAGTATGACGATGACGTACCTTCTGTGTAAGTCCAAAGGGGAAACCAATTCA 1781
Db 1360 -GluGluGlnGluAspGluGln-ProProTriPAlaSerPro-Ser-----ProThrSer 1376
QY 1782 CGCGTTAGGAACAGTGACTCG-----GCCCCACCTCGTGNACACAAA 1826
Db 1377 ArgGlnSerProAlaSerProProProLeuGlyGlyGlyAlaProValArgThrAlaLys 1396
QY 1827 GCCTCGGACGACCTTGGAGAGGGCCACATGACACACACAGATGCGATCCTTGGGACCT 1886
Db 1397 Ala-----GluArgArgHis----- 1401
QY 1887 GAATCTATCACCCAGGAATCTCAAACTCCCTTTGGCCCTGAACGAGGCGCAGATAAGGAA 1946
Db 1402 -----GlnGluArgLeuArgValGlnSerProGluProProAlaProGluArg 1417
QY 1947 CAGCTGGGCCACTTTT-TTGAAGGCCAATGTGGAGGAAGGGAGCAGCCGCTTTGG 2005
Db 1418 AlaLeuSerProAlaGluLeuArgAla-----LeuGluAlaGluLysArgAlaLeuTriP 1435
QY 2006 GAGAAGATCTCAAGGATCCAGACTCTCATTCCTTCTCTGCGCCAGTGAATTTGGTCTC 2065
Db 1436 ---ArgAlaAlaArgMetLysSerLeuGluGlnAspAlaLeuArgAlaGlnMetValLeu 1454
QY 2066 TCCGAGCTTTGG-----GGGACT----- 2083
Db 1455 SerArgSerGlnGluGlyArgGlyThrArgGlyProLeuGluArgLeuAlaGluAlaPro 1474
QY 2084 ---CCTTCCTTGACCGCTAATAGACCCGACTGGAGTCTCTC----- 2122
Db 1475 SerProAlaProThrProSerProThrProValGluAspLeuGlyProGlnThrSerThr 1494
QY 2123 -----TCTCTCCATCCCTCTCCT 2140
Db 1495 SerProGlyArgLeuSerProAspPheAlaGluLeuArgSerLeuGluProSerPro 1514
QY 2141 CTGCC 2146
Db 1515 SerPro 1516
```

Search completed: March 21, 2003, 12:59:28
Job time : 108.681 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 08:38:27 ; Search time 52.8072 Seconds
(without alignments)
10910.938 Million cell updates/sec

Title: US-09-502-945-5

Perfect score: 3989

Sequence: 1 cctggccggtgcggtgcg.....tatttccgcgttaaaaaa 2162

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09502945/runat_14032003_101057_19087/app_query.fasta_1.10979
-DB=A_Geneseq_101002 -QPMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLW=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptio -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09502945_@CGN_1_1_365 @runat_14032003_101057_19087 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEFTIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	3196.5	80.1	652	20	AA107094
2	3196.5	80.1	652	20	AA107050
3	2188	54.9	481	20	AA107098
4	2168	54.3	521	20	AA107099
5	2046	51.3	403	20	AA107095
6	1557.5	39.0	328	20	AA107097
7	1175.5	29.5	268	20	AA138440
8	435	10.9	120	21	AA133238
9	418	10.5	493	22	ABB61924
10	417	10.5	98	20	AA138390
11	406	10.2	95	20	AA138437
12	271	6.8	1535	23	AA1087934
13	267.5	6.7	1037	22	ASG22366
14	267	6.7	928	23	AAE21718
15	263	6.6	856	21	AA101383
16	256	6.6	2037	21	AA153753
17	256	6.4	763	20	AA104741
18	256	6.4	1005	20	AA104731
19	256	6.4	1373	20	AA104730
20	256	6.4	2000	20	AA104732
21	256	6.4	2070	20	AA104733
22	254	6.4	1111	23	AA1087918
23	250.5	6.3	1881	20	AA124025
24	249	6.2	1552	22	AA108793
25	249	6.2	1554	22	ASG07288
26	237	5.9	46	22	ABB38777
27	237	5.9	46	22	ABB23814
28	237	5.9	46	22	AA159416
29	237	5.9	46	22	AA11974
30	237	5.9	46	22	AA119386
31	237	5.9	46	22	AA12249
32	237	5.9	46	23	ASG41789
33	236	5.9	1745	23	AB157253
34	234.5	5.9	767	23	AA1084269
35	234.5	5.9	767	23	AA1084328
36	234	5.9	1526	22	AA107977
37	232.5	5.8	767	21	AA122138
38	232.5	5.8	817	19	AA148101
39	232.5	5.8	817	20	AA130137
40	232.5	5.8	849	19	AA148102
41	232	5.8	45	22	ABB38778
42	232	5.8	45	22	ABB23815
43	232	5.8	45	22	AA159417
44	232	5.8	45	22	AA11975
45	232	5.8	45	22	AA119387

ALIGNMENTS

RESULT 1
AA107094
ID AA107094 standard; Protein; 652 AA.
XX
AC AA107094;
XX
DT 02-JUL-1999 (first entry)
XX
DE Colon cancer associated antigen precursor sequence.

XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.

XX PF 15-JUL-1998; 98WO-US14679.
 XX PR 22-JUN-1998; 98US-0102322.
 XX PR 17-JUL-1997; 97US-0896164.
 XX PR 10-OCT-1997; 97US-0061599.
 XX PR 10-OCT-1997; 97US-0061765.
 XX PR 10-OCT-1997; 97US-0948705.
 XX PR 11-OCT-1997; 97GB-0021697.
 XX PA (LUDW-) LUDWIG INST CANCER RES.
 XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
 XX PI Tureci O;
 XX DR WPI; 1999-132448/11.
 XX PT New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 XX PS Disclosure; Page 656-658; 787pp; English.
 XX CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 XX SQ Sequence 652 AA;
 Alignment Scores:
 Pred. No.: 2,77e-268 Length: 652
 Score: 3196.50 Matches: 626
 Percent Similarity: 96.02% Conservative: 1
 Best Local Similarity: 95.87% Mismatches: 0
 Query Match: 80.13% Indels: 26
 DB: 20 Gaps: 1
 US-09-502-945-5 (1-2162) x AAY07094 (1-652)
 QY 97 ATGGACCAAGAGTGGCCGAGAAATCCGGCATAGGTGGATTTCTGATTGAAAATGAT 156
 Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
 QY 157 GCAGAGAAGGACTATCTCTATGATGTGCTGGGAATGTACCAACAGACCATGGAGCTGGCC 216
 Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
 QY 217 GTGCTGTGGAGACCTGAAGCTGGTCAATCAATGAACCCAGCCGCTGCCCTCTGTTGAT 276
 Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
 QY 277 GCATTGGCCGCTGATCCCACTGAAGCACAGGTGGAAATATGATCATCTGACCCCGG 336
 Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 QY 337 CGCTCCAGGAAGCTGAAGAGGTGGCTGTGGACCGCTCTGCACCCCGAAGGCGCTGGGCTG 396
 Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
 QY 397 AGTGTGCGTGGGCTCGAGTTTGGCTGTGGGCTCTTCATCTCCCACTCATCAAAAGGC 456
 Db 460 sGluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProIleGlyLysValva 480

Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
 QY 457 GGTTCAGGCAGACAGCGTCCGGCTCCAGGTAGGGACGAGATCGTCCGGATCAATGGATAT 516
 Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 QY 517 TCATCTCTCTCTGATCCCATGAGGAGGTATCAACCTCATTCGACCAAGAAACTGTG 576
 Db 141 SerIleSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
 QY 577 TCATCAAAAGTGAGACACATCGCCCTGATCCCGTGAAGCTCTCTCATGAGCCCTC 636
 Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
 QY 637 ACTTGGCAGTATGTGATCAGTTTGTGTCGGAATCTGGGGCGCTGCGAGCGACCTGGGC 696
 Db 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200
 QY 697 TCCCTGGAAATCGGGAACAAAGGAGAGAGGTCTTTCATGAGCTGTAGCTCCCA 756
 Db 201 SerProGlyAsnArgGluAsnLysGluLysValPheIleSerLeuValGlySerArg 220
 QY 757 GGCTTGGCTGAGCAGCATTTCCAGCGCCCATCCAGAACCTGGCATCTTTATCAGCCAT 816
 Db 221 GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
 QY 817 GTGAAACCTGGTCCCTGCTGCTGAGGTGGATGGAGATAGGGACAGCATTTGTCGAA 876
 Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
 QY 877 GTCATATGGGCTGACCTCTCTACCTGGATACAGAGAGCTGTAATGTGCTGAAAAAT 936
 Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
 QY 937 AGCCGAGCCTGACCATCTCCATTGAGCTGACGCTGGCGGGAGCTGTTTCATGACAGAC 996
 Db 281 SerArgSerLeuThrIleSerIleValAlaAlaAlaGlyArgGluLeuPheMetThrAsp 300
 QY 997 CGGAGCGGCTGGCAGAGCGCGGCGAGCTGAGCTCAGCGGAGGAGCTTCTCATGTCAG 1056
 Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
 QY 1057 AACCGGCTGGCGATGGAGTCCAAACAGATCTCCAGGACGAGGAGGATGGAGCGCAA 1116
 Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGluGlnGluMetGluArgGln 340
 QY 1117 AGGAGAAAAGAAATTCGCCAAGAGGACAGAGAGAAATGAGAGATACCGAAGAGAGATG 1176
 Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
 QY 1177 GAACAGATTGTAGAGGAGGAAGAAGTTTAAAGAACAATGGGAAGAAGCTGGGGCTCA 1236
 Db 361 GluGlnIleValGluGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380
 QY 1237 AAGAACAGCTACTCTTTCCTTAAACCATCCTGCTGAGTACACCCAGTACCCCTTCGC 1296
 Db 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 QY 1297 AAGCCA----- 1302
 Db 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspAspLeuAspGly 420
 QY 1303 -----AAGTATTTCCGGAATAATATGAGAAGGCTTTCACCCCTA 1341
 Db 421 GlyThrGluGluGlnGlyGluGln-AspPheArgLysTyrGluGluGlyPheAspProTy 440
 QY 1342 CTCTATCTTCCACCCAGACAGATCATGGGAAGGATGTCCGGCTCTCTAGCATCAGAA 1401
 Db 440 rSerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLy 460
 QY 1402 GGAGGGATCTTACAGCTGCCCTGGAGCGGTGTGGACTCCCTCCCATTCGGAGGTGGT 1461
 Db 460 sGluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProIleGlyLysValva 480

QY 1462 CGTTTCTGCTGTATGAGCGGGAGCTGCTGAGCGGCATGGTGCATGTCTAAGGGGA 1521
 Db 480 lvalserAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAs 500
 QY 1522 CGAGATCATGGCAATCAACGGCAAGATTGTGACAGCTACACCCCTGGCTGAGGCTGACGC 1581
 Db 500 pglulleMetaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAl 520
 QY 1582 TGCCCTCGAGAGGCTCGAATCAGGGGGGAGCTGGATCGACCTGTGTGGTTCGGCTGTG 1641
 Db 520 aAlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValValAlaValCy 540
 QY 1642 CCCCCCAAGAGTATGAGTACGATGACCTGCTTCTGCTGAAGTCCAAAGGGGAACCA 1701
 Db 540 sProProLysGlnTyrAspAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnG 560
 QY 1702 AATTCACCGTGTAGAAACAGTACGCTCCGGCCCGCCACCTCGTGAACACAAAGCCCTCGGAC 1761
 Db 560 nIleHisAlaLeuGlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgTh 580
 QY 1762 CAGCCTTGAGAGGCCACATGACACACACAGATGGCATCCTTGGGACCTGAATCTATC 1821
 Db 580 rSerLeuGluArgGlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSe 600
 QY 1822 ACCCAGGAATCTCAAACTCCCTTTGGCCCTGAACCCAGGCCAGATAAGAACAGCTCGGG 1881
 Db 600 rProArgAsnLeuLysLeuProLeuAlaLeuAsnGlnGlyGlnIleArgAsnSerSerG 620
 QY 1882 CCACCTTTTTGAAGGCCAATGTGGAGGAAAGGAGCAGCCAGCCGCTTTGGGAGAGATCT 1941
 Db 620 yHisPheGluGlyGlnCysGlyGlyLysGlyAlaAlaSerArgLeuGlyGluAspLe 640
 QY 1942 CAAGGATCCAGACTCTATCTCTTCTTCTGCGCCAG 1978
 Db 640 ulysaspProaspSerHisSerPheProLeuAlaGln 652
 RESULT 2
 AAY07050
 ID AAY07050 standard; Protein: 652 AA.
 XX AC AAY07050;
 XX DT 02-JUL-1999 (first entry)
 XX DE Renal cancer associated antigen precursor sequence.
 XX KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.
 XX OS Homo sapiens.
 XX PN W09904265-A2.
 XX PD 28-JAN-1999.
 XX PF 15-JUL-1998; 98WO-US14679.
 XX PR 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 PI Pfeundsschuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 XX WPI: 1999-132448/11.

XX

PT New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers

XX PS Disclosure; Page 436-438; 787pp; English.

XX CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.

XX SQ Sequence 652 AA;

Alignment Scores:
 Pred. No.: 2,77e-268 Length: 652
 Score: 3196.50 Matches: 626
 Percent Similarity: 96.02% Conservative: 1
 Best Local Similarity: 95.87% Mismatches: 0
 Query Match: 80.13% Indels: 26
 DB: 20 Gaps: 1

US-09-502-945-5 (1-2162) x AAY07050 (1-652)

QY 97 ATGGACCGAAAAGTGGCCGAGAAATCCGGCATAAGCTGGATTTCCTGATTGAAATGAT 156

Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20

QY 157 GCAGAGAGGACTATCTCTATGATGTCTGCGAATGTACCACAGACCATGGAGCTGGCC 216

Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40

QY 217 GTGCTCTGGGAGACCTGAAGCTGTCATCAATGAACCCAGCGCTGCTCTGTTTGTAT 276

Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60

QY 277 GCCATTCCGGCGCTGATCCCACTGAAGCAGCAGGTGGAATATGATCAGCTGACCCCGCG 336

Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80

QY 337 CGCTCCAGGAAGCTGAAGAGGTCGCTGTGGACCGCTGTGACCCCGGAGCGCTCGGCGCTG 396

Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100

QY 397 AGTGTGCTGGTGGCGCTGGAGTTGGCTGTGGCTCTTCATCTCCCACTCATCAAGGC 456

Db 101 SerValArgGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120

QY 457 GGTCCAGCAGACAGCTCGGCTCCAGGTAGGGGACGAGATCGTCCGGATCAATGGATAT 516

Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspIleValArgIleAsnGlyTyr 140

QY 517 TCCATCTCCTCCTGTACCCCATGAGGAGGTGCATCAACCTCATTTGACCAAGAACTGTG 576

Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160

QY 577 TCCATCAAGTGAGACACATCGGCGCTGATCCCGTGAAAGCTCTCTCTGATGAGCCCTC 636

Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180

QY 637 ACTTGGCAGTATGTGGATCAGTTTGTCTCGGAATCTGGGGCGCTGGCAGGACGCTGGGC 696

Db 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200

QY 697 TCCTCGGAATCGGAAACAGAGAGAGAGGTCTTCATCAGCCTGGTAGGCTCCCGA 756
Db 201 SerProGlyAsnArgGluAsnLysGluLysValPheLeuValGlySerArg 220
QY 757 GGCCTTGGCTGAGCATTTCCAGCGCCGCCCATCCAGAGAGCTGGCATCTTTATCAGCCAT 816
Db 221 GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTGAACCTGGCTCCCTGCTCTGCTGAGTGGGATGGAGATAGGGAGCAGATTCGAA 876
Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCATGGCTGCTCTTCTTAACCTGGATCACAGGAGGCTGTAATGCTGCTGAAAT 936
Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluValAlaAsnValLeuLysAsn 280
QY 937 AGCCGAGCCTGACCATCTCCATTGTAGCTGCAGCTGGCCGGAGCTGTTTCATGACAG 996
Db 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGGAGCGCTGGCAGAGCGCGGAGCGTGAGCTGCAGCGCAGAGCTTCTCATGCGAG 1056
Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
QY 1057 AAGCGCTGGCGATGGAGTCCCAACAGATCCCTCCAGGAGCAGCAGAGATGGAGCGCAA 1116
Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln 340
QY 1117 AGGAGAAAGAAATGCCAGAGGAGCAGCAGAGGAGAAATGAGAGATACCGGAGGAGATG 1176
Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
QY 1177 GRACATTTGTAGAGGAGGAGAGAGATTTAGAACCAATGGAGAACAGACTGGGGCTCA 1236
Db 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluAspTrpGlySer 380
QY 1237 AAGGAACAGCTACTTGTGCTTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGC 1296
Db 381 LysGluGlnLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCCA----- 1302
Db 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
QY 1303 -----AAGTGATTTCCGGAATATGAGAAAGGCTTTGACCCCTA 1341
Db 421 GlyThrGluGluGlnGlyGluGln-AspPheArgLysTyrGluGluGlyPheAspProty 440
QY 1342 CTCTATGTTACCCAGAGCAGATCATGGGGAAGGATGTCCTGGCTCCTACGCATCAAGAA 1401
Db 440 rSerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLy 460
QY 1402 GGAGGGATCTTAGACCTGGCCCTGGAGCGCTGTGGACTCCCCATTTGGGAGGTGT 1461
Db 460 sGluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProIleGlyLysValVa 480
QY 1462 CGTTTCTGCTGTATGAGCGGGAGCTGCTGAGCGGCATGTGGCATTTGTAAGGGGA 1521
Db 480 lValSerAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAs 500
QY 1522 CGAGATCATGGCAATCAACGGCAAGATTGTACAGACTACACCCCTGGCTGAGGCTGACGC 1581
Db 500 pGluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAl 520
QY 1582 TGGCCTCAGAGGCTGGGAATCAGGCGGGGACTGGATGCACCTGTGTGTTGCCGTCTG 1641
Db 520 aAlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValValAlaValCy 540
QY 1642 CCCCCAAAGAGTATCACCATGAGCTGACCTTCTTCTCAAGTCCCAAGGGAACCA 1701
Db 540 sProLysGluTyrAspAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnGln 560

QY 1702 AATCAGCGCTTAGGAACACAGTGTAGCTCGGCCCCACCTCCTGTGAACACAAAGCCTCGAC 1761
Db 560 nIleHisAlaLeuGlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgTh 580
QY 1762 CAGCCTTGAGAGAGCCACATGACACACACAGATGCATCCTTGGACCTGAATCTATC 1821
Db 580 rSerLeuGluArgGlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSe 600
QY 1822 ACCAGGAATCTCAACTCCCTTTGGCCCTGAACCCAGGCGCCAGATAAGGAACAGCTCGG 1881
Db 600 rProArgAsnLeuLysLeuProLeuAlaLeuAsnGlnGlnIleArgAsnSerSerG1 620
QY 1882 CCACCTTTTGAAGGCCAATCTGGAGGAAGGAGCAGCCGCTTTGGGAGAGATCT 1941
Db 620 yHisPhePheGluGlyGlnCysGlyGlyLysGlyAlaAlaSerArgLeuGlyGluAspLe 640
QY 1942 CAAGATCCACACTCTCATCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1978
Db 640 uLysAspProAspSerHisSerPheProLeuAlaGln 652
RESULT 3
AAY07098
ID AAY07098 standard; Protein; 481 AA.
XX
AC AAY07098;
XX
DT 02-JUL-1999 (first entry)
XX
DE Colon cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Disclosure; Page 664-665; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.

XX SQ Sequence 481 AA;

Alignment Scores:

Pred. No.: 7,41e-181 Length: 481
 Score: 2188.00 Matches: 438
 Percent Similarity: 91.06% Conservative: 0
 Best Local Similarity: 91.06% Mismatches: 1
 Query Match: 54.85% Indels: 43
 DB: 20 Gaps: 1

US-09-502-945-5 (1-2162) x AAY07098 (1-481)

Qy 97 ATGGACGAAAAGTGGCCGCGAGAAATCCGGCATAAGGTGGATTTCTGATTGAAATGAT 156
 Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
 Qy 157 GCAGAGAGGACTATCTATGATGCTGCGCAATGTACCACACACCATGGACGTGGCC 216
 Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
 Qy 217 GTGCTCGTGGGAGACCTGAAGCTGCTCATCAATGAACCCAGCGCTGCTCTGTTTGTAT 276
 Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
 Qy 277 GCCATTGGCCGCTGATCCCACTGAAGCACCAGGTGGATATGATCAGCTGACCCCGCGG 336
 Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 Qy 337 CGCTCCAGGAGCTGAAGGAGGTGGCTGTGGACCGTCTGCACCCGCGAGGCTCGGCCGTG 396
 Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
 Qy 397 AGTGTGCTGTGGCTGGAGTTTGGCTGGTGGCTTTCATCTCCACCTCATCAAGAGGC 456
 Db 101 SerValArgGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
 Qy 457 GGTACGAGCAGACAGCTCGGGCTCCAGGTAGGGGACGAGATCGTCCGGATCAATGATAT 516
 Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 Qy 517 TCCATCTCTCTGTACCCATGAGAGGTCAATCAACCTCAATTCGAACCAAGAAACTGTG 576
 Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
 Qy 577 TCCATCAAGTGAGACATCGGCTGATCCCGTGAAGAGCTCTCCTGATGAGCCCTC 636
 Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
 Qy 637 ACTTGGCAGTATGTGGATCAGTTTGTGCGAATCTGGGGCGTGGAGGAGCGCTGGCC 696
 Db 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200
 Qy 697 TCCCTTGAATTCGGGAAACAGAGAGAAGGTCTTCATCAGCGCTGGTAGGCTCCCGA 756
 Db 201 SerProGlyAsnArgGluAsnLysGluLysLysValPheIleSerLeuValGlySerArg 220
 Qy 757 GGCCTTGGCTGCAGCATTTCCAGCGGCCCATCCAGAGCCCTGCGATCTTTATCAGCCAT 816
 Db 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
 Qy 817 GTGAACCTGGCTCCCTGCTGCTGAGGTGGGATTGGAGATAGGGGACGAGATTGTCGAA 876
 Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
 Qy 877 GTCAAATGGCGTCTTCTTAACCTGGATCACAAGGAGGTGTAAATGTGCTGAAAAAT 936
 Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280

Qy 937 AGCGCAGCCTGACATCTCCATTGTAGCTCAGCTGCGCGGAGCTGCTTTCATCAGAC 996
 Db 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
 Qy 997 CGGAGCGGCTGGCAGAGCGCGCAGCGTGCAGCGCAGGAGCTTCTCATGCAG 1056
 Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
 Qy 1057 AAGCGGCTGGCGATGGAGTCCAAAGATCTCCAGGAGCAGCAGGAGATGGAGCGCAA 1116
 Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGluGlnGluMetGluArgGln 340
 Qy 1117 AGGAGAAAAGAAATGCCACAGAGCAGCAGGAGAAATGAGATACCGGAAGGAGATG 1176
 Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
 Qy 1177 GAACAGATTGTAGAGGAGGAGAGAGTAAAGCAATGGGAGAGACTGGGGCTCA 1236
 Db 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380
 Qy 1237 AAGGAACAGCTACTTCTTCCCTAAACCATCACTGCTGAGGTACACCGATCCCTTCGC 1296
 Db 381 LysGluGlnLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 Qy 1297 AAGCCAAAGTG----- 1307
 Db 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspAspLeuAspGly 420
 Qy 1307 ----- 1307
 Db 421 GlyThrGluGluGlnGlyGluGlnProGlnGluMetLeuLysArgMetValValTyrGln 440
 Qy 1308 -----ATTTCGGGAAATATGAGGAGGCTTTGACCCCTACTCTATGT 1349
 Db 441 AspSerIleGlnAspLysIleSerGlyAsnMetArgLysAlaLeuThrProThrLeuCys 460
 Qy 1350 TCAACCCAGACAGCATGATGGGAGGATGTCGCGCTCCTACGATCAAGAGAGGAGAT 1409
 Db 461 SerProGlnSerArgSerTrpGlyArgMetSerGlySerTyrAlaSerArgArgAsp 480
 Qy 1410 CCT 1412
 Db 481 Pro 481
 RESULT 4
 AAY07099
 ID AAY07099 standard; Protein; 521 AA.
 AC AAY07099;
 XX
 XX 02-JUL-1999 (first entry)
 XX
 DE Colon cancer associated antigen precursor sequence.
 XX
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.
 XX
 OS Homo sapiens.
 PN WO9904265-A2.
 XX
 PD 28-JAN-1999.
 XX
 PF 15-JUL-1998; 98WO-US14679.
 XX
 PR 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX

PA (LUDW-) LUDWIG INST CANCER RES.
 XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 XX WPI: 1999-132448/11.
 XX New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 XX
 PS Disclosure: Page 666-667; 787pp; English.
 XX
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 XX
 SQ Sequence 521 AA;
 Alignment Scores:
 Pred. No.: 4, 18e-179 Length: 521
 Score: 2168.00 Matches: 438
 Percent Similarity: 84.07% Conservative: 0
 Best Local Similarity: 84.07% Mismatches: 1
 Query Match: 54.35% Indels: 83
 DB: 20 Gaps: 1
 US-09-502-945-5 (1-2162) x AAY07099 (1-521)
 QY 97 ATGACCGAAAGTGGCCGAGAAATCCGCGATAGTGGATTCTCTGATTGAAATGAT 156
 Db 1 MCAAPARGLYSVALAARGGLUPHEARGHISLYSVALASPPHELEULEGLUASNASP 20
 QY 157 GCAGAGAAGACTATCTCTATGATGTGCTGCGGAATGTACCACAGACCACCTGGACGTGGCC 216
 Db 21 ALAAGLYSASPTYRLEUTYRASPVALLAARGMETYRHISGLNTHRMETASPVALLA 40
 QY 217 GTGCTGTGGAGACCTGAAGCTGGTCATCAATGAACCCAGCCGCTGCGCTCTGTTGAT 276
 Db 41 VALLEUVALGLYASPLEULYSLEUVALILEASNGLUPROSERARGLEUPROLEUPHEASP 60
 QY 277 GCATTCGCGCGCTGATCCCACTGAGCACCAGTGGATATGATCAGTGCACCCCGG 336
 Db 61 ALAILEARGPROLEULEPROLEULYSGLNVALGLUTYRASPVALLAARGMETYRHISGLNTHRMETASPVALLA 80
 QY 337 CGCTCCAGGAAGCTGAAGAGGTGCGTCTGACCGCTGACCCCGAAGCCCTCGCGCTG 396
 Db 81 ARGSERARGLYSLEULYSGLUVALARGLEUASPARGLEUHLISPROGLUGLYLEUGLYLEU 100
 QY 397 AGTGTGCGTGGTGGCTGGATTGGCTGTGGCTCTTTCATCTCCACCTCATCAAGGC 456
 Db 101 SERVALARGGLYGLYLEUGLUPHEGLYCYSGLYLEUPHEILESERHISLEULLELYSGLY 120
 QY 457 GGTGAGCAGACAGCTGCGGCTCCAGGTAGGGAGCAGATCTCGGATCAATGATAT 516
 Db 121 GLYCINALAASPSERVALGLYLEUGLINVALGLYASPVALLAARGMETYRHISGLNTHRMETASPVALLA 140
 QY 517 TCCATCTCTCTCTACCATGAGGAGTTCATCAACTTCATTCGACCCAGAAACCTGG 576
 Db 141 SERILESERCYSTRHISGLUGLUVALILEASNLEULEARGTHRILYSLYSTHRVAL 160

577 TCCATCAAGTGAACACATCGCGCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTC 636
 Db 161 SERILELYSVALARGHISILEGLYLEUILEPROVALYSSERSERPROASPGLEUPROLEU 180
 QY 637 ACTTGGCAGTATGTGATCAGTTTGTGCGAATCTGGGGCGTGGCAGGCGCTGGGC 696
 Db 181 THRTRPGLNTRYVALASPGINPHEVALSERGLUSERGLYGLYVALARGGLYSERLEUGLY 200
 QY 697 TCCCTCGAAATCGGAAAAACAAGGAGAAGCTTTCATCAGCCTGGTAGGCTCCCGA 756
 Db 201 SERPROGLYASNARGLEUASNLYSLYSLYVALPHEILESERLEUVALGLYSERARG 220
 QY 757 GGCCTTGGCTGCAGCATTTCCAGCGGCCCCCATCCAGAGCCTGGCATCTTATCAGCCAT 816
 Db 221 GLYLEUGLYCYSERILESERSERGLYPROFLEGLNLYSPROGLYILEPHEILESERHIS 240
 QY 817 GTGAAACCTGCTCCCTGCTCTGCTGAGTGGGATTGAGATGAGGACCGACCATGTCTGAA 876
 Db 241 VALLYSPROGLYSERLEUSERALAGLUVALGLYLEUGLULEGLYASPGINLEVALGLU 260
 QY 877 GTCAATGCGCTGACTTCTTAACTGGATCACAAGGAGGCTGTAATGTGCTGAAAT 936
 Db 261 VALASNGLYVALASPPHESERASNLEUASNLYSLYSLYGLUVALASNVALLEULYSASN 280
 QY 937 AGCCGACCTGACCATCTCCATTGTAGCTGCAGCTGCCGCGGAGCTGTTTCATCAGACAC 996
 Db 281 SERARGSERLEUTHRILESERILEVALALAALAAGLYARGGLULEUPHEMETHRASP 300
 QY 997 CGGAGCGCTGGCAGAGCGCGCTGAGCTGCAGCGCAGCGAGGAGCTTCTCATGCGAG 1056
 Db 301 ARGGLUARGLEUALAAGLUALAARGGLNARGGLULEUGLINGLULEULEUMETGLN 320
 QY 1057 AAGCGCTGGCGATGGAGTCCCAACAGATCCTCCAGGACGACGAGAGTAGGAGCGGCAA 1116
 Db 321 LYSARGLEUALAMETGLUSERASNLYSILELEUGLINGLUGLINGLUMETGLUARGLN 340
 QY 1117 AGGAGAAAGAAATGCCCCAGAGGCACGAGGAATAGAGATACCGGAAGGAGATG 1176
 Db 341 ARGARGLYSGLUILEALAGLNLYSALAALAAGLUASNGLUARGTYRARGLYSGLU 360
 QY 1177 GAACAGATTGTAGAGGAGGAAGAGAAAGTTTAAAGCAATGGGAAGAGACTGGGGTCA 1236
 Db 361 GLUGLILEVALGLUGLUGLULYSPhelysLysGlnTrpGLUGLUASPTPGLYSER 380
 QY 1237 AAGAACAGCTACTCTTCCTTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGC 1296
 Db 381 LYSGLUGLULEULEUPROLYSTHRIETHRALAGLUVALHISPROVALPROLEUARG 400
 QY 1297 RAGCCAAAGTG----- 1307
 Db 401 LYSPROLYSTYRASPGINGLYVALGLUPROGLULEUGLUPROALAASPSLEUASPGLY 420
 QY 1307 ----- 1307
 Db 421 GLYTHRGLUGLUGLUGLINTHRPHECYSPROSERPROGLINPROPROARGGLYPRO 440
 QY 1307 ----- 1307
 Db 441 GLYVALSERTHRILESERLYSVALMETVALHISGLINGLUPROASNPHILETYRARG 460
 QY 1307 ----- 1307
 Db 461 PROALAVALLYSSERGLEUVALLEUPROGLINGLUMETLEULYSARGMETVALTYRGLN 480
 QY 1308 -----ATTTCCGGAATATGAGAAAGCTTTTCCACCCCTACTCTATGT 1349
 Db 481 ASPSERILEGLINASPGLYSILESERGLYASNMETARGLYSALALETURHLEUTHRLEUCYS 500
 QY 1350 TCACCCAGCAGCAGATCATGCGGAAGGATGTCGGCTCCTACGCATCAAGAGGAGGAT 1409
 Db 501 SERPROGLINSEARGSERTRPPOLYARGMETSERGLYSERTYRALASERARGARGASP 520
 QY 1410 CCT 1412

```

Db 521 Pro 521
RESULT 5
ID AAY07095 standard; Protein; 403 AA.
XX AC AAY07095;
XX DT 02-JUL-1999 (first entry)
XX DE Colon cancer associated antigen precursor sequence.
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX OS Homo sapiens.
XX PN W09904265-A2.
XX PD 28-JAN-1999.
XX PF 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
XX PR 10-OCT-1997; 97US-0061599.
XX PR 10-OCT-1997; 97US-0061765.
XX PR 10-OCT-1997; 97US-0948705.
XX PR 11-OCT-1997; 97GB-0021697.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX DR WPI; 1999-132448/11.
XX PT New isolated cancer associated nucleic acids and polypeptides -
XX PT isolated using sera from cancer patients, used to develop products
XX PT for the diagnosis, monitoring or treatment of cancers
XX PS Disclosure; Page 659-660; 787pp; English.
XX CC The invention relates to a method for diagnosing a disorder characterised
XX CC by expression of a human cancer associated antigen precursor coded for by
XX CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX CC biological sample isolated from a subject with an agent that specifically
XX CC binds to the NAM, an expression product or a fragment of an expression
XX CC product complexed with an HLA molecule; and (b) determining the
XX CC interaction between the agent and the NAM or the expression product as a
XX CC determination of the disorder. The products and methods can be used in
XX CC the diagnosis, monitoring, research, or treatment of conditions
XX CC characterised by the expression of various cancer associated antigens.
XX CC The invention provides nucleic acid sequences and encoded polypeptides
XX CC which are cancer associated antigen precursors expressed in human breast
XX CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX CC lung cancer.
XX SQ Sequence 403 AA;

Alignment Scores:
Pred. No.: 1.42e-168 Length: 403
Score: 2046.00 Matches: 403
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 51.29% Indels: 0
DB: 20 Gaps: 0

US-09-502-945-5 (1-2162) x AAY07095 (1-403)

```

```

QY 97 ATGACCGAAGAGTGGCCGCGAGATTCGCGCATAGAGTGGATTCTCTGATTGAAATGAT 156
Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
QY 157 GCAGAGAAGGACTATCTCTATGATGTCTGCGAATGTACCACAGACCATGGACGTGGCC 216
Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
QY 217 GTGCTCGTGGAGACCTGAAGCTGGTCATCAATGAACCCAGCCGCTCTGCTCTGTTGAT 276
Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
QY 277 GCATTCCGGCGGTGATCCCACTCAAGCAGCAGGTGGGAATATCATCAGCTGACCCCGG 336
Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
QY 337 CGCTCCAGGAAGCTGAAGAGAGTCCGCTCTGACCGCTGTGCACCCGAGCCCTCGGCGTG 396
Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
QY 397 AGTGTGCTGGTGGCTGGAGTTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGGC 456
Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
QY 457 GGTCAAGCAGACAGCTCGGCTCCAGGTAGGGAGGAGATCGCTCCGGATCAATGATAT 516
Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
QY 517 TCCATCTCCCTCTGTACCCATGAGAGGTTCATCAACTTCATCGAACCAAGAAACTGTG 576
Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
QY 577 TCCATCAAGTGAGACACATCGGCTGATCCCGTGAAGAGCTCTCTGTATGAGCCCTC 636
Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
QY 637 ACTTGGCAGTATGTGATCAGTTTGTGCGAATCTGGGGCTGGGGAGCAGCCCTGGGC 696
Db 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlySerLeuGly 200
QY 697 TCCCTCGAATCGGGAACAAAGAGAGAGTCTTCATCAGCTGTGTAGGCTCCCGA 756
Db 201 SerProGlyAsnArgGluAsnLysGluLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCTTGGCTGAGCATTTCCAGCGGCCCATCCAGAGCCTGGCATCTTTATCAGCCAT 816
Db 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTAAACCTCGCTCCCTCTCTGCTGAGTGGGATTTGAGATAGGGAGCAGATTGTGAA 876
Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCAATGGCTCGCATTTCTTAACCTGGATTCACAAGAGGCTGTAATGTCTGTAATAAT 936
Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
QY 937 AGCGAGCCTGACCATCTCCATTTAGCTGACGTGGCGGAGCTGTTTCATCAGACAG 996
Db 281 SerArgSerLeuThrIleSerIleValAlaAlaAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGGAGCGGCTGGCAGAGCGCGGCTGAGCTGACGCGGAGGAGCTTCTCATGTCAG 1056
Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuLeuMetGln 320
QY 1057 AAGCGCTGGCGATGGAGTCCAAAGATCTCCAGGAGCAGGAGGATGGAGCGGCAA 1116
Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnGlnGlnGln 340
QY 1117 AGGAGAAAGAAATTCGCCAGAGGAGCAGAGAGAAATCAGAGATACCGGAGGAGATG 1176
Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
QY 1177 GACAGATTCTAGAGGAGGAGAGAGATTAAAGCAATGGGAAGAGACTGGGCTCA 1236

```

```

|||||
Db 361 GluGlnIleValGluGluGluGluLysPheLysGlnTrpGluGluAspTrpGlySer 380
QY 1237 AAGGAACAGCTACTTCCCTAAACCATCATCTGCTGAGGTACACCCAGTACCCCTTCGC 1296
|||||
Db 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCCAAAG 1305
Db 401 LysProLys 403
|||||
RESULT 6
RAY07097
ID AAY07097 standard; Protein: 328 AA.
AC AAY07097;
XX 02-JUL-1999 (first entry)
XX Colon cancer associated antigen precursor sequence.
DE Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX Homo sapiens.
OS
PN W09904265-A2.
XX 28-JAN-1999.
XX 15-JUL-1998; 98WO-US14679.
XX 22-JUN-1998; 98US-0102322.
PR 17-JUN-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI; 1999-132448/11.
XX
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Disclosure; Page 662-663; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX Sequence 328 AA;
SQ
Alignment Scores:
Pred. No.: 3.09e-126 Length: 328

```

```

Score: 1557.50 Matches: 302
Percent Similarity: 92.10% Conservative: 1
Best Local Similarity: 91.79% Mismatches: 0
Query Match: 39.04% Indels: 26
DB: 20 Gaps: 1
US-09-502-945-5 (1-2162) x AAY07097 (1-328)
QY 1069 ATGGAGTCCCAACAAGATCCTCCAGAGCAGCAGAGATGAGCGGCAAGAGAGAAAGAA 1128
|||||
Db 1 MetGluSerAsnLysIleLeuGlnGluGlnMetGluArgGlnArgArgLysGlu 20
QY 1129 ATTCCCCAGAGGCAGCAGAGGAAATCAGATACCGGAGGAGATGGACACATGTGA 1188
|||||
Db 21 IleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMetGluGlnIleVal 40
QY 1189 GAGGAGGAAGAGAAGTTTAAAGAACAATGGGAAGAAGACTGGGGCTCAAGGAACAGCTA 1248
|||||
Db 41 GluGluGluLysPheLysLysGlnTrpGluLysTrpGlySerLysGluGlnLeu 60
QY 1249 CTCTTGCTTAAACCATCATCTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCA----- 1302
|||||
Db 61 LeuLeuProLysThrIleThrAlaGluValHisProValProLeuArgLysProLysTyr 80
QY 1302 ----- 1302
Db 81 AspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGlyGlyThrGluGlu 100
QY 1303 -----AAGTGATTTCCGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTTTCAAC 1353
|||||
Db 101 GlnGlyGluGln-AspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheTh 120
QY 1354 CCCAGAGCAGATCATGCGGGAAGGATGTCGGCTCTCTAGCGATCAAGAAGGAGGATCCTT 1413
Db 120 rProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLe 140
QY 1414 AGACTGCGCTTGGGAAGCGGTGGTGGCTCCCATGGGAAGTGGTCTGCTGTGT 1473
Db 140 uAspLeuAlaLeuGluGlyGlyValAspSerProIleGlyLysValValSerAlaVal 160
QY 1474 GTATGAGCGGGAGCTGCTGAGCGCATGTGTGGCATTTGTAAAGGGGACGAGATCATGGC 1533
Db 160 lTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAl 180
QY 1534 AATCAACGGCAAGATGTGACAGACTACACCTGGCTGAGGCTGACGCTGCCCTGCAGAA 1593
Db 180 aIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAlaLeuGlnLy 200
QY 1594 GGCCTGGATCAGGGCGGGAGTGGATCGACTGTGTGGTTCGCTGCGCCCAAGGA 1653
Db 200 sAlaTrpAsnGlnGlyGlyAspTrpIleAspLeuValValAlaValCysProProLysGl 220
QY 1654 GTATGACGATGAGCTGACCTTCTTGTGAGTCCAAAGGGGAAACCAATTCACGCGTT 1713
Db 220 uTyrAspAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnGlnIleHisAlaLe 240
QY 1714 AGGAACAGTGTGAGTCCGGCCCTCCCTCGTGAACACAAAGCTCGGACCAAGCTTGAGAG 1773
Db 240 uGlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgThrSerLeuGluArg 260
QY 1774 AGGCCACATGACACACACCATGTCCTTGGGACCTGAATCTATCACCACCAAGTCT 1833
Db 260 gGlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSerProArgAsnLe 280
QY 1834 CAACCTCCCTTTGGCCCTTGAACACCGGCGCAGATGAAGAACAGCTCGGCGCCACTTTTGA 1893
Db 280 uLysLeuProLeuAlaLeuAsnGlnGlyIleArgAsnSerSerGlyHisPhePheGl 300
QY 1894 AGGCCAATGTGGAGAAAGGAGGAGCAGCCGCTTTGGGAGAGATCTCAGGATCCAGA 1953
Db 300 uGlyGlnCysGlyGlyLysGlyAlaAlaSerArgLeuGlyGluAspLeuLysAspProAs 320
QY 1954 CTCATTCCTTCTCTGCGCCAG 1978

```

Db 320 pSerHisSerPheProLeuAlaGln 328
 RESULT 7
 AAY38440
 ID AAY38440 standard; Protein; 268 AA.
 AC AAY38440;
 XX
 DT 18-OCT-1999 (first entry)
 XX
 DE Human secreted protein encoded by gene No. 5.
 XX
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 PN WO9935158-A1.
 XX
 PD 15-JUL-1999.
 XX
 PF 06-JAN-1999; 99WO-US00108.
 XX
 PR 07-JAN-1998; 98US-0070704.
 PR 07-JAN-1998; 98US-0070657.
 PR 07-JAN-1998; 98US-0070658.
 PR 07-JAN-1998; 98US-0070692.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Duan RD, Ebner R, Lafleur DW, NI J;
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
 XX
 DR WPI; 1999-444190/37.
 DR N-PSDB; AA206223.
 XX
 PT New isolated human genes and the secreted polypeptides they encode
 XX
 PS Disclosure; Page 206-207; 227pp; English.
 CC
 CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AA206210) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 36 novel genes and their fragments (nucleic
 CC acid sequences: AA206219-206263; amino acid sequences AAY38386-Y38498)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 36
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AA206219 for described uses).
 XX
 SQ Sequence 268 AA;
 Alignment Scores:
 Pred. No.: 3,87e-93 Length: 268
 Score: 1175.50 Matches: 236
 Percent Similarity: 90.11% Conservative: 1
 Best Local Similarity: 89.73% Mismatches: 0
 Query Match: 29.47% Indels: 26
 DB: 20 Gaps: 1
 US-09-502-945-5 (1-2162) x AAY38440 (1-268)

QY 964 GCTGCAGCTGGCGGAGCTGTTTCATGACAGACCGGAGCGGCTGCAGAGCGCGCAG 1023
 Db 6 AlaAlaAlaGlyArgGluLeuPheMetThrAspArgGluArgLeuAlaGluAlaArgGln 25
 QY 1024 CGTGAGCTGCAGCGGAGGAGCTTCTCATGCAGAACCGGCTGGCGATGGAGTCCACAG 1083
 Db 26 ArgGluLeuGlnArgGlnGluLeuMetGlnLysArgLeuAlaMetGluSerAsnLys 45
 QY 1084 ATCCTCCAGGAGCAGGAGGATGGAGCGCAAGAGAGAAAATAATGCCAGAGGCA 1143
 Db 46 IleLeuGlnGlnGlnGlnGlnMetGluArgGlnArgArgLysGluIleAlaGlnLysAla 65
 QY 1144 GCAGAGAAATAGAGATACCGGAAGAGATGGAACAGATGTAGAGAGAGAGAGAG 1203
 Db 66 AlaGluGluAsnGluArgTyrArgLysGluMetGluGlnIleValGluGluGluLys 85
 QY 1204 TTTAACAACAATGGGAACAAGCTGGGCTCAAGAGACACACTACTCTTGCCTAAAC 1263
 Db 86 PheLysLysGlnTrpGluAspTrpGlySerLysGluGlnLeuLeuLeuProLysThr 105
 QY 1264 ATCCTGCTGAGTACACCCAGTACCCCTTCGCAAGCCA----- 1302
 Db 106 IleThrAlaGluValHisProValProLeuArgLysProLysTyrAspGlnGlyValGlu 125
 QY 1303 -----AAGTGA 1308
 Db 126 ProGluLeuGluProAlaAspAspLeuAspGlyGlyThrGluGluGlnGlyGluGln-As 145
 QY 1309 TTTCCGGAATATGAGGAGGCTTTGACCCCTACTCTATGTTCCACCCAGACAGATCAT 1368
 Db 145 pPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGluGlnIleMe 165
 QY 1369 GGGGAAGGATGTCGGCTCTACGATCAAGAGAGGAGGATCTTAGACCTGGCCCTGGA 1428
 Db 165 TGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeuAspLeuAlaLeuGlu 185
 QY 1429 AGCGGTGTGGACTCCCGCTCCCGCTGGAGGTGTGCTGTTCTGTGTGTATGAGCGGGAGC 1488
 Db 185 uGlyGlyValAspSerProIleGlyLysValValSerAlaValTyrGluArgGlyAl 205
 QY 1489 TGCTAGCGGATGTTGGCATTTGAAAGGGAGGAGATCATGGCAATCAACGGCAAGAT 1548
 Db 205 alaGluArgHisGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLysIle 225
 QY 1549 TGTGACAGACTACACCTCGCTGAGCTGACGCTGCCCTGCAGAGAGGCTTGAATCAGG 1608
 Db 225 eValThrAspTyrThrLeuAlaGluAlaAspAlaAlaLeuGlnLysAlaTrpAsnGlnGlu 245
 QY 1609 CGGGGACTGCAGCTGCTGTTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
 Db 245 yGlyAspTrpIleAspLeuValAlaValCysProLysGluTyrAspGluLeu 265
 QY 1669 GACCTTC 1675
 Db 265 urThrPhe 267
 RESULT 8
 AAB53238
 ID AAB53238 standard; Protein; 120 AA.
 XX
 AC AAB53238;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human colon cancer antigen protein sequence SEQ ID NO:778.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antineoplastic; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;

KW infectious disease; cardiovascular disorder.

XX Homo sapiens.

XX PN W020005351-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05883.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA,* Ruben SM;

XX DR WPI; 2000-587534/55.

XX DR N-PSDB; AAC97995.

XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer

XX PS Claim 11; Page 1339; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and
CC vulnerable, nephrotoxic, antineoplastic and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.

XX SQ Sequence 120 AA;

Alignment Scores:
Pred. No.: 4, 62e-29 Length: 120
Score: 435.00 Matches: 97
Percent Similarity: 80.65% Conservative: 3
Best Local Similarity: 78.23% Mismatches: 14
Query Match: 10.90% Indels: 10
DB: 21 Gaps: 4

US-09-502-945-5 (1-2162) x AAB53238 (1-120)

Qy 1196 AAGAGAGTTTAAGA---AGCAATCGGAAGAGATGGGCTCAAGAGACAGTACTCT 1252

Db 6 GluArgHisLeuArgValThrGlnGlyLeuGlyThrGlyAlaPheLeuGlyGly----- 23

Qy 1253 TCCCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTGGATTTC 1312

Db 24 ---LeuArgProValLeu-----GlnProArgGlnGlyGln---AspPhe 36

Qy 1313 CGGAATATGAGGAGGCTTTGACCCCTACTCTATGTTACACCCAGACGATCATGGGG 1372

Db 37 ArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGluGlnIleMetGly 56

Qy 1373 AAGATGTCGGCTCCTACGCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAAGGC 1432

Db 57 LysAspValArgLeuArgIleLysLysGluGlySerLeuAspLeuAlaLeuGlyGly 76

Qy 1433 GGTGTGGACTCCCATTTGGGAGGTGGTCTGTTCTGCTGTGTATGACGGGGAGCTGCT 1492

Db 77 GlyValAspSer***IleGlyLysValValSerAlaValTyrGluArgGlyAlaAla 96

Qy 1493 GAGCGCATGGTGCATTTGTGAAGGGGACGAGATCATGCAATCAACGGCAAGATTGTG 1552
Db 97 GluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLysIleVal 116

Qy 1553 ACAGACTACACC 1564

Db 117 ThrAspTyrThr 120

RESULT 9

ABB61924

ID ABB61924 standard; Protein; 493 AA.

XX AC ABB61924;

XX DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 12564.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX PN W0200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL06027.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions

XX PS Disclosure; SEQ ID NO 12564; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS7072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 493 AA;

Alignment Scores:
Pred. No.: 2, 66e-27 Length: 493
Score: 418.00 Matches: 128
Percent Similarity: 45.89% Conservative: 62
Best Local Similarity: 30.92% Mismatches: 157
Query Match: 10.48% Indels: 67
DB: 22 Gaps: 12

US-09-502-945-5 (1-2162) x ABB61924 (1-493)

Qy 340 TCCAGGAGCTGAGGAGGTGCGCTCTGGACGCTGCACCCCGAAGGCCCTC----- 390

Db 58 SerSerArgLeuArgValLeuArgLeuValArgProHisGlnArgArgLeuValGly 77

```
QY 391 -----GGCTGAGTGTGGTGGCTGGCTGGAGTTGGCTGT 426
Db 78 GlyProGluArgGlySerThrTyrglyPhehrValArgGlyGlyArgGlyHisGlyThr 97
QY 427 GGGCTTTCATCTCCCACTCATCAAGAGCGGTGAGGACAGAGCGTGGGCTCCAGGTA 486
Db 98 GlyPhePheValSerHisValGluHisGlyGlyGluAlaHisLeuLysGlyLeuArgile 117
QY 487 GGGACGAGATCGCCGATCAATGATATTCATCTCTCTGTACCATCAGAGATC 546
Db 118 GlyAspGlnIleLeuArgileAsnGlyPheArgLeuAspAspAlaValHisLysGluPhe 137
QY 547 ATCAACCTCATTCGAACCAAGAAAGTGTCCATCAAGTGAGACACATCGGCTGATC 606
Db 138 IleGlnLeuValAlaGlyGlnAspArgValThrLeuLysValArgGlyValGlyMetLeu 157
QY 607 CCCGTGAAGCTCTCCGTGAGCCCTCACTTGGCAGTATGTGGATCAGTTGTGTGCG 666
Db 158 ProValArgAspLeuProGluGluArgLeuSerTrpSerValValLys-----LeuPro 175
QY 667 GAATCTGGGGCGTCCGAGGAGCGCTGGCTCCCTGGAATCGGGAACCAAGAGAG 726
Db 176 SerValSerGlyThrProSerGlnSerSerPheLysGlyGluArgArgGlyAlaSerArg 195
QY 727 AAGGTCTTTCATCAGCCTG---GTAGGCTCCCGA---GGCCTTGGCTGCAGCATTTCCAGC 780
Db 196 AspileSerValValLeuHisValAlaProArgThrLysLeuGlyLysLysLys 215
QY 781 GGCCCATCCAGACCTCGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGTCTGCT 840
Db 216 GlyProGluTrpLysProGlyIlePheValGlnPheThrLysAspArgSerValAlaArg 235
QY 841 GAGGTGGATGGAGATGAGGAGGACAGATTCGNAAGTCATGCGTCCGCTCTCTTAAC 900
Db 236 GluAlaGlyLeuArgProGlyAspGlnIleLeuSerValAsnSerIleAspPheSerAsp 255
QY 901 CTGGATCAACAGAGGTGTAAATGTCTGAAATAATGCCAGCCTGACCATCTCCATT 960
Db 256 ValLeuPheSerGluAlaValAlaValMetLysSerSerLysLeuAspMetValVal 275
QY 961 GTACTGCGAGTGGCGGGAGCTGTCATGACAGCGGGAGCGGTGGCAGAGCGCGG 1020
Db 276 ArgThrAlaAlaGlyCysAspLeuPheProGlyGlu-SerSerGlyTyraSerSerAl 295
QY 1021 CAGCGTGAGTGCAGCGCAGGAGCTTCTCATGACAGAGCGGTGGCGATGAGTCCAAC 1080
Db 295 aSer---SerValThrGlyAspGlnSerProCys-----TrpAlaAspAl 309
QY 1081 AAGATCTCTCCAGAGCAGCAGGATGGAGCGCAAGAGAAAGAAATTTGCCAGAG 1140
Db 309 aLysSerLysArg-LeuThr-----AlaValArgGluGluSerGlyAlaGlyGlyG 326
QY 1141 GCAGCAGAGGAAATGACAGATACCGGAAGGAGATGGAACAGATGTAGAGGAGGAAG 1200
Db 326 IyGlyCysGlyLeuSerSerAlaProGlyAlaGlySerProAsnTrpSerClnGlyValG 346
QY 1201 AAGTTTAAAGCAATGGGAAGACATG----- 1229
Db 346 luValHisLysGlnMetAsnLysThrIleLysLeuThrGluAsnGlyThrSerIleA 366
QY 1230 -----GGCTCAAGAGAA 1242
Db 366 snAsnThrTyriAlaSerThrGlySerSerValSerSerGlySerGlySerThrGlys 386
QY 1243 CAGTACTCTTTCCTAAACACCATCCTGCTGAGGTACACCGATCCCTTCGCAAGCCA 1302
Db 386 erGlyThrSerGlyArgSerGlnGlnSerGlnSerAsnProSerAsnProSerArgAsn- 405
QY 1303 AAGTGATTTCCGGAATATGAGGAGGCT---TTGACCCCTACT----- 1343
Db 406 -----SerThrThrMetLysArgSerHisLeuArgProValAsnSerAlaGlySerG 423
```

QY 1344 -----CTATGTTCAACCCAGCAGATCATGGGAAGG-----ATGTCGGGCTCTC 1389
Db 423 lyIleGlyLeuSerSerGlySerAlaGlySerAlaGlySerAlaGlySerGlySerG 443
QY 1390 ACGCATCAAGAAGAGGAGGATCCTTAGACCTGGCCCT 1425
Db 443 lySer-ArgSerGlyGlyValIleAlaProAlaPro 454

RESULT 10
AAY38390
ID AAY38390 standard; Protein; 98 AA.
XX AAY38390;
XX 30-SEP-1999 (first entry)
XX Human secreted protein encoded by gene No. 5.
DE Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; gastritis; neoplasm;
KW endocrine; metabolism; regulation; malabsorption;
XX Homo sapiens.
OS WO9935158-A1.
PN 15-JUL-1999.
PD 06-JAN-1999; 99WO-0500108.
XX 07-JAN-1998; 98US-0070704.
PR 07-JAN-1998; 98US-0070657.
PR 07-JAN-1998; 98US-0070558.
PR 07-JAN-1998; 98US-0070892.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
XX WPI; 1999-444190/37.
DR N-PSDB; AA206223.
XX New isolated human genes and the secreted polypeptides they encode
PT Claim 11; Page 178-179; 227pp; English.
XX This sequence represents a secreted human protein encoded by the gene
CC clone detailed in the descriptor line. The gene can be used to generate
CC fusion proteins by linking to the gene to a human immunoglobulin Fc
CC portion (e.g. AA206210) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 36 novel genes and their fragments (nucleic
CC acid sequences: AA206219-206263; amino acid sequences AAY38386-X38498)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 36
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AA206219 for described uses).
XX Sequence 98 AA;
SQ

Alignment Scores:
Pred. No.: 1 53e-27 Length: 98
Score: 417.00 Matches: 88
Percent Similarity: 83.81% Conservative: 0

XX	This sequence represents a secreted human protein encoded by the gene			
CC	clone detailed in the descriptor line. The gene can be used to generate			
CC	fusion proteins by linking to the gene to a human immunoglobulin Fc			
CC	portion (e.g. AA206210) for increasing the stability of the fused			
CC	protein as compared to the human protein only.			
CC	The invention relates to 36 novel genes and their fragments (nucleic			
CC	acid sequences: AA206219-206263; amino acid sequences AAY38386-Y38498)			
CC	which are useful for preventing, treating or ameliorating medical			
CC	conditions e.g. by protein or gene therapy. Also, pathological			
CC	conditions can be diagnosed by determining the amount of the new			
CC	polypeptides in a sample or by determining the presence of mutations in			
CC	the new polynucleotides. Specific uses are described for each of the 36			
CC	polynucleotides, based on which tissues they are most highly expressed in			
CC	(see AA206219 for described uses).			
XX				
SQ	Sequence	95 AA;		
Alignment Scores:				
Pred. No.:	1.36e-26	Length:	95	
Score:	406.00	Matches:	90	
Percent Similarity:	60.13%	Conservative:	2	
Best Local Similarity:	58.82%	Mismatches:	1	
Query Match:	10.18%	Indels:	60	
DB:	20	Gaps:	1	
US-09-502-945-5 (1-2162) x AAY38437 (1-95)				
QY	1678	GCTGAAGTCCAAAGGGAAACCAAAATTCACGCGTTAGGAAACAGTGAGTCCGGCCCCA	1737	
DB	3	SerGluValGlnLysGlyLysProAsnSer-ProLeuGlyAsnSerGluLeuArgProH1	22	
QY	1738	CCTCGTGAACACAAAGCCTCGGACCGCTTGAGAGAGGCCACATGACACACACAGATG	1797	
DB	22	sLeuValAsnThrLysProArgThrSerLeuGluArgGlyHis-----	36	
QY	1798	GCATCCTTTGGGACCTGAATCTATCACCAGCAATCTCAAACTCCCTTTGGCCCTGAACCA	1857	
DB	36	-----	36	
QY	1858	GGGCCAGATAAGGAACAGCTCGGCCACTTTTGAAGGCCAATGTGGAGAAAGGGAGC	1917	
DB	36	-----	36	
QY	1918	AGCCAGCGCTTTGGGAGAAGATCTCAAGAGATCCAGACTCTCATCTTCTCTGGCCCCA	1977	
DB	37	-----Thr-IleProPheLeuTrpProS 44		
QY	1978	GTGAATTTGGTCTCTCCAGCTTTGGGGAGCTCTTCTTGAACCCCTAATAAGACCCAC	2037	
DB	44	erGluPheGlyLeuSerGlnLeuTrpGlyThrProSerLeuAsnProAsnLysThrProL 64		
QY	2038	TGGAGTCTCTCTCTCCATCCCTCTCTGCGCTCTGCTCTAATTTGCTGCCAGGATTG	2097	
DB	64	euGluSerLeuSerLeuHisProSerProLeuProSerAlaLeuAlaAlaArgIleV 84		
QY	2098	TCACTCCAAACCTTACTCTGAGCTCATTAATAAAA 2132		
DB	84	alThrProAsnLeuThrLeuSerSerLeuLys 95		
RESULT 12				
ID	AAU87934	standard; Protein; 1535 AA.		
XX	AAU87934;			
XX	05-JUN-2002	(first entry)		
DT	Human protein containing PDZ domain.			
DE	Human; PDZ domain; WW domain; rat; cow; mouse; fruitfly; protein therapy;			
KW	gene therapy; PDZ-mediated disease; inward potassium channel; WBP;			
KW	dimer inhibitor peptide; carboxylate binding loop.			


```
QY 1344 CTATGTTCCACCCAGAGCAGATCATGGGAGGATGTCGGCTCTACGCATCAAGAAGG 1403
Db 1215 ProCysSerProSerGlyGln-GlnProProSerProSerProAspGluLeuProAl 1234
QY 1404 AGGGATCCTTAGA----- 1416
Db 1234 aAsnValLysGlnAlaTyArgAlaPheAlaValProThrSerHisProProGluAs 1254
QY 1417 -----CCTGGCCCTGGAAG----- 1431
Db 1254 pAlaProAlaGlnProProThrProGlyProAlaAlaSerProGluGlnLeuSerPheAr 1274
QY 1432 -----CGGTGTGGATCTCCCCATTTGGGAAGT-----GG 1460
Db 1274 gGluArgGlnLysTyPheGluLeuGluValArgValProGlnAlaGluGlyProProLy 1294
QY 1461 TCGTTTCTGCTGTATGAGCGGGAGCTGCTGACGGCATGGTGGCATTTGTAAGGGG 1520
Db 1294 sArgValSerLeuVal-----G 1300
QY 1521 ACGAGATCATGCGCAATCAACGCGCAAGATTGTGACAGACTACACCCTGCTGAGGCTGACG 1580
Db 1300 yAlaAspAspLeuArgLysMetGlnGlu----- 1309
QY 1581 CTGCCCCGAGAAGGCCTTGGATCAGGCGGGGACTGGATCGACCTTGTGTGCGCT 1640
Db 1310 -----GluGluAlaArgLysLeuGlnGln----- 1317
QY 1641 GCGCCCCAAGGAGTATGACGATGAGCTGCTTCTGTGAGTCCAAAGGGGAACCC 1700
Db 1318 -----LysArgAlaGlnMetLeuArgGluAlaAlaGluAlaGlyAl 1331
QY 1701 AAATTCACGCTTAGGAACAGTGTGAGCTCCGCGCCACCTGTGAACACAAAGCCTCGGA 1760
Db 1331 aGluAlaArgLeuAlaLeuAspGlyGluThrLeuGlyGluGluGlnGluAspGluGl 1351
QY 1761 CCAGCCTTGAGAGGCCACATGACACACAGATGCGATCTTGGGAGCTGAATCTAT 1820
Db 1351 nProProTAlaSerProSerProThrSerArgGlnSerProAlaSerProProProLe 1371
QY 1821 C-----ACCCAGGAATCTCAACTCC 1841
Db 1371 uGlyGlyAlaProValArgThrAlaLysAlaGluArgHisGlnGluArgLeuArgVa 1391
QY 1842 CTTTGGCCTTGACACAGGCGCAGATAGGAACAGCTCGGCGCCACTTTT-TTGAAGGCCAA 1900
Db 1391 lGlnSerProGluProProAlaProGluArgAlaLeuSerProAlaGluLeuArgAla-- 1410
QY 1901 TGTGGAGAAAGGAGCAGCCAGCCGTTTGGGAGAAGATCTCAAGATCCAGACTCTAT 1960
Db 1411 -----LeuGluAlaGluLysArgAlaLeuTrp---ArgAlaAlaArgMetLysSerLeuGl 1428
QY 1961 TCCTTTCTCTGCGCCAGTCAATTTGCTCTCTCCAGCTTTGG-----GGGACT-- 2009
Db 1428 uGlnAspAlaLeuArgAlaGlnMetValLeuSerArgSerGlnGluGlyArgGlyThrAr 1448
QY 2010 -----CCTTCCTTGAACCCCTAATAAGACCCC 2035
Db 1448 gGlyProLeuGluArgLeuAlaGluAlaProSerProAlaProThrProSerProThrPr 1468
QY 2036 ACTGGAGTCTCTC----- 2048
Db 1468 oValGluAspLeuGlyProGlnThrSerThrSerProGlyArgLeuSerProAspPheAl 1488
QY 2049 -----TCTCTCCATCCCTCTCTCTGCC 2072
Db 1488 aGluGluLeuArgSerLeuGluProSerProSerPro 1500
RESULT 13
ABG22366
ID ABG22366 standard; Protein; 1037 AA.
XX
AC ABG22366;
```

```
XX 18-FEB-2002 (first entry)
DT
DE
XX
XX Novel human diagnostic protein #22357.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS86553.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
XX
XX Claim 20; SEQ ID No 52725; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(CC) (II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG0010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
```

XX Sequence 1037 AA;

Alignment Scores:

Pred. No.:	4.28e-14	Length:	1037
Score:	267.50	Matches:	157
Percent Similarity:	35.09%	Conservative:	89
Best Local Similarity:	22.40%	Mismatches:	220
Query Match:	6.71%	Indels:	235
DB:	22	Gaps:	32

US-09-502-945-5 (1-2162) x ABG22366 (1-1037)

QY 183 GCTGGGAATGTACACACAGAC-----CATGGACGTGGCGCTGCTGGGGAGACT 233

Db 83 SerAlaIleValProGlyAspGlyValAlaGlyGlyGlyThrGlyAspGlyGlyPro 102

QY 234 GAAGCTGTTCATCAATGAACCCAGCCGCTCGCTCTGTTTGTATGCCATTCGCCCTGAT 293

Db 103 GlyAlaProTyr----- 106
QY 294 CCACCTGAGCAGCAGCTGGATATGATCAGCTGACCTGACCCCGCCCTCAGGAAGCTGAA 353
Db 107 -----ProGlyAlaValAlaAspMetGluGluLeuThrIleTrpGluGlnHis 122
QY 354 GGAGGTGGCTGTGGACCGCTCTCACCCGAGAGCGCTCGCCTGAGCTGCGGTGGCGCT 413
Db 123 ThrAlaThrLeuSerLysAspProArgArg-GlyPheGlyIleAlaIleSerGlyGlyArg 142
QY 414 GGAGTTT---GGCTGTGGCTCTTCATCTCCACCTCATCAAGCGCGTCAGGCAGACAG 470
Db 142 gAspArgProGlyGlySerMetValValSerAspValProGlyGlyProAlaGluG1 162
QY 471 CGTCGGCTCCAGGTAGGGAGCAGATCGTCGGATCAATGGATATTCATCTCCTCTGCT 530
Db 162 y---ArgLeuGlnThrGlyAspHisIleValMetValAsnGlyValSerMetGluAsnAl 181
QY 531 TACCCATGAGGAGGTCACTCAACCTCATTCGAACC---AAGAAACTGTGTCCATCAAGT 587
Db 181 aThrSerAlaPheAlaIleGlnIleLeuLysThrCysThrLysMetAlaAsnIleThrVa 201
QY 588 G-----AGACATATCGGCTGATCCCGTGAAGAGCTCTCT----- 624
Db 201 LlysArgProArgArgIleHisLeuProAlaThrLysAlaSerProSerSerProGlyArg 221
QY 625 -----GATGACCCCTCACTTGGCGATGTGGATCAG----- 657
Db 221 gGlnAspSerAspGluAspAspGlyProGlnArgValGluGluValAspGlnGlyArgG1 241
QY 658 -----TTTGTGTGGATCTGG----- 675
Db 241 yTyraSpGlyAspSerSerSerGlyArgSerTrpAspGluArgSerArgArgPr 261
QY 676 -----GGCTGCGAGGAGCAGCTGGCTCCCTCGAAATCGGAAACAAAGGAGAGAA 728
Db 261 oArgProGlyArgArgGlyArgAlaGlySerHisGlyArgSerProGlyGly----- 279
QY 729 GGTCTTCATCAGCTGTAGCTCCGAGGCTTGGCTGCGAGCATTTCCAGCGG----- 783
Db 280 -----GlySerGluAlaAsnGlyLeuAlaLeuValSerGlyPheLy 293
QY 784 -----OCCATCCAG-----AAGCCT----- 798
Db 293 sArgLeuProArgGlnAspValGlnMetLysProValLysSerValLeuValLysArgArg 313
QY 799 -----GGCATCTTTATCAGCCATGTGAACC 824
Db 313 gAspSerGluGluPheGlyValLysLeuGlySerGlnIlePheIleLysHisIleThrAs 333
QY 825 TGGCTCCCTGTCTGCT---GAGGTGGGATTGGAGATAGGGACAGATTGTGCAAGTCAA 881
Db 333 pSerGlyLeuAlaAlaArgHisArgGlyLeuGlnGluGlyAspLeuIleLeuGlnIleAs 353
QY 882 TGGCTCGACTCTCTAACCCTGGATCACAAGAGGCTGTAATGTGCTGAAATATAGCCG 941
Db 353 nGlyValSerSerGlnAsnLeuSerLeuAsnAspThrArgLeuIleGluLysSerGl 373
QY 942 CAGCTGACCATCTCATTAGCTGACGCTGGCGGAGCTGTTCATGACAGAC----- 996
Db 373 uGly---LysLeuSerLeuValLeuArgAspArgGlyGlnPheLeuValAsnIlePr 392
QY 997 -----CGGGAGCG 1004
Db 392 oProAlaValSerAspSerSerProLeuGluAspIleSerAspLeuAlaSerGl 412
QY 1005 GCTGCGCAGAGCG-----CGGCAGCGTGAAGTGCAGCGCGCAGGA 1043
Db 412 uLeuSerGlnAlaProProSerHisIleProProProProArgHisAlaGlnArgSerPr 432
QY 1044 GCTTCTCATGCAAGAGCGCTGGCGATGAGTCCAAAGATCTCTCCAGGAGCAGCAGGA 1103
Db 432 oGluAlaSerGlnThrAspSerProValGluSerProArgLeu----- 446

QY 1104 GATGAGCGGCAAGAGGAAAGAAATTTGCCAGAAAGCAGAGAGGAAATATGAGATA 1163
Db 447 -----ArgArgGluSerSerValAspSerArgThrIleSerGluProAspGluG1 463
QY 1164 CCGAAGAGAGATGAACAGATTTGTAGAGGAGGAGGAAGATTTTAAGAAGCAATGGGAAGA 1223
Db 463 nArgSerGlu----- 466
QY 1224 AGACTGGGCTCAAAGGAACAGCTACTCTTGCTTAAACCATCATCTGCTGAGGTACACCC 1283
Db 467 -----LeuProArgGluSerSerTyraSpIleTyra 477
QY 1284 AGTACCCCTTCGCAAGCAAAAGTATTTCCGGAATATATGAGGAAGGCTTTGACCCCTACT 1343
Db 477 gVal-Pro---SerSerGlnSer-----MetGluAspArgGlyTyraSerPro--- 491
QY 1344 CTATGTTTACCCCAAGCAGATCATGGGAAGGATGTCCGGCTCTACGACATCAAGAGG 1403
Db 492 -----AspThrArgValValArgPheLeuLysG 501
QY 1404 AGGATCTTAGACCTGGCCCTGGAAGCGGTGTGGACTCCCCCATTTGGAAAGGTGTGCTG 1463
Db 501 LysSerIleGlyLeuArgLeuAlaGlyGlyAsnAsp-----ValGly---IlePheV 518
QY 1464 TTTCTGTGTATGACGGGGAGCTCTCTGAGCGGCATGGTGGCATTTGTGAAGGGGAGC 1523
Db 518 a1SerGlyVal---GlnAlaGlySerProAlaAspGlyGlnGlyIleGlnGluGlyAspG 537
QY 1524 AGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTGGCTGGCTGAGCGCTG 1583
Db 537 InIleLeuGlnValAsnAspValProPheGlnAsnLeuThrArgGluGluAla----- 554
QY 1584 CCCTGCAGAGGCTGGAATCAGGGCGGAGCTGGATCGACCTGTGTGTGGCTGTGCTGCC 1643
Db 555 -----ValGlnPheLeuLeuGlyLeuProp 563
QY 1644 CCCCAAGAGATATGAC-----GATGAGCTGACCTTCTTCTGCTGA 1682
Db 563 roGlyGluGluMetGluLeuValThrGlnArgLysGlnAspIlePheTrpLysMetValG 583
QY 1683 AGTCCAAAGGGGAAACAAATTCACCGCTTAGGAACACATGAGCTCCGCGCCACCTCG 1742
Db 583 InSerArgValGlyAspSerPheTyraIleArgThrHisPheGluLeuGluPro----- 600
QY 1743 TGAACAAAGCTCTCGGACCGCTT-----GAGAGAGCCACATGACACACACC----- 1792
Db 601 -----SerProProSerGlyLeuGlyPheThrArgGlyAspValPheHisValLeuA 618
QY 1793 --AGATGGCATCTTGGGACCTGAATCTATCACCAGGAATCTCAAACTCCCTTTGGCCC 1850
Db 618 spThrLeuHisPro----- 622
QY 1851 TGAACAGGCGCCAGATAAGGACAGCTCGGCCACCTTTTGAAGGCCAATGTGGAGGAA 1910
Db 623 --GlyProGlyGlnSerHisAlaArgGlyGlyHisTrpLeu----- 635
QY 1911 AGGAGCAGCAGCCGCTTTGGGAGAGATCTCAAGGATCCAGACTCTCATTTCTTCTCT 1969
Db 636 -----AlaValArgMetGlyArgAspLeuArgGluGlnGluArgGlyIleIlePro 652
RESULT 14
AAE21718
ID AAE21718 standard; Protein; 928 AA.
XX AC AAE21718;
XX AC
XX AC
DT 16-JUL-2002 (first entry)
XX DE Human PKIN-13 protein.
XX DE
KW Human; kinase; enzyme; PKIN-13 protein; immune system disorder; anaemia;
KW acquired immune deficiency syndrome; thymic hypoplasia; Crohn's disease;

Db 340 SerProArgLeu-----ArgArgGluSerSerValasp 350
Qy 1135 CAGAGGCGAGAGGAGAAATGACAGATACCGGAGAGAGATGAGAGAG 1194
Db 351 SerArgThrIleSerGluProAspGluArgSerGlu----- 363
Qy 1195 GAAGAGAAGTTAAGAACATGGGAAGACACTGGGCTCAAGGACACACTACTCTTG 1254
Db 364 -----Leu 364
Qy 1255 CTAAGACCATCACTGCTAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATTCCG 1314
Db 365 ProArgGluSerSerArgAspIleArgValPro-----SerSerGlnSer----- 380
Qy 1315 GAATATGAGGAGGCTTTGACCCCTACTCTATGTTACCCAGCAGACATCATGGGAA 1374
Db 381 -MetGluAspArgGlyTyrSerPro----- 388
Qy 1375 GGATGCGGCTCCTACGCATCAAGAGGAGGATCCTTAGACCTGCCCTGGAAGCGG 1434
Db 389 -AspThrArgValValArgPheLeuIleGlySerIleGlyLeuArgLeuAlaGlyG 408
Qy 1435 TGTGACTCCGCCATGGAAGGAGTGTCTTCTGCTGTATGAGCGGGAGCTGCTGA 1494
Db 408 YasnAsp-----ValGly---IlePheValSerGlyVal---GlnAlaGlySerProAl 424
Qy 1495 CGGCATGCTGCATTTGAAGGGGAGGAGATCATGCAATCAACGGCAAGATGTGAC 1554
Db 424 aAspGlyGlnGlyIleGlnGlyArgPheGlnIleLeuGlnValAsnAspValProPheG 444
Qy 1555 AGACTACACCTGGCTGAGGCTGAGCTGCCCTGCAGAGGCTGGAATCAGCGCGGGA 1614
Db 444 nAsnLeuThrArgGluGluAla----- 451
Qy 1615 CTGATGACGACCTGTGTGTCGCTGCCGCCCAAGAGGATGATGAC----- 1660
Db 452 ---ValGlnPheLeuLeuGlyLeuProGlyGluGluMetGluLeuValThrGlnAr 470
Qy 1661 -----GATGAGCTGACCTTCTGTGTAAGTCAAAAGGGGAAACCAATTCACGGTT 1713
Db 470 gLysGlnAspIlePheTrpLysMetValGlnSerArgValGlyAspSerPheTyrIleAr 490
Qy 1714 AGAAACAGTGCAGCTCGCGGCCACCTCGTCAACACAAAGCTCGGACACGCTT----- 1768
Db 490 gThrHisPheGluLeuGluPro-----SerProSerGlyLeuGlyPh 505
Qy 1769 -GAGAGGCGCACATGACACACACC-----AGATGGCATCCTTGGGACCTGAATCTATC 1821
Db 505 eThrArgGlyAspValPheHisValLeuAspThrLeuHisPro----- 519
Qy 1822 ACCAGGATCTCAACTCCCTTTGGCCCTGAACGAGGCGCAGATAGGAACACGCTCGG 1881
Db 520 -----GlyProGlyGlnSerHisAlaArgGlyG 529
Qy 1882 CCACCTTTTGAAGGCCAATGTGGAGGAAGGAGGACGACCGCTTGGGAGAGATCT 1941
Db 529 yHisTrpLeu-----AlaValArgMetGlyArgAspLe 540
Qy 1942 CAAGATCCAGACTCTCATCTTCTTCCT 1969
Db 540 uArgGluGlnGluArgGlyIleIlePro 549
RESULT 15
AAB01383
ID AAB01383 standard; Protein; 856 AA.
XX AC
XX AC
XX AC
DT 20-OCT-2000 (first entry)
XX AC
DE Neuron-associated protein.
XX AC
KW Neuron associated protein; NEUAP; neurological disorder; epilepsy;

ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW Alzheimer's disease; Pick's disease; Huntington's disease;
KW dementia; Parkinson's disease; demyelinating disease; meningitis;
KW prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;
KW cerebral palsy; muscular dystrophy; central nervous system; CNS;
KW peripheral nervous system; PNS; myopathy; schizophrenia;
KW actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW AIDS; Addison's disease; adult respiratory distress syndrome;
KW allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW Werner syndrome; trauma; human.
XX Homo sapiens.
XX OS
PH Domain
FT 1..65
FT /label= PD2 domain signature
FT Modified-site 18
FT /note= "Potential glycosylation site"
FT Modified-site 20
FT /note= "Potential phosphorylation site"
FT Modified-site 69
FT /note= "Potential phosphorylation site"
FT Modified-site 71
FT /note= "Potential phosphorylation site"
FT Modified-site 73
FT /note= "Potential phosphorylation site"
FT Modified-site 118
FT /note= "Potential phosphorylation site"
FT Modified-site 119
FT /note= "Potential phosphorylation site"
FT Domain 136..218
FT /label= PD2 domain signature
FT Modified-site 144..147
FT /note= "Glycosaminoglycan attachment site"
FT Modified-site 199
FT /note= "Potential glycosylation site"
FT Modified-site 238
FT /note= "Potential phosphorylation site"
FT Modified-site 253
FT /note= "Potential phosphorylation site"
FT Modified-site 263
FT /note= "Potential phosphorylation site"
FT Domain 269..349
FT /label= PD2 domain signature
FT Modified-site 284
FT /note= "Potential phosphorylation site"
FT Modified-site 296
FT /note= "Potential phosphorylation site"
FT Modified-site 337
FT /note= "Potential phosphorylation site"
FT Modified-site 341
FT /note= "Potential phosphorylation site"
FT Modified-site 369
FT /note= "Potential glycosylation site"
FT Modified-site 370
FT /note= "Potential phosphorylation site"
FT Modified-site 386
FT /note= "Potential phosphorylation site"
FT Modified-site 389
FT /note= "Potential glycosylation site"
FT Modified-site 390
FT /note= "Potential phosphorylation site"
FT Modified-site 414
FT /note= "Potential phosphorylation site"
FT Domain 415..497
FT /label= PD2 domain signature
FT Modified-site 475
FT /note= "Potential phosphorylation site"
FT Domain 511..592
FT /label= PD2 domain signature

FT Modified-site 531 /note= "Potential glycosylation site"
 FT Modified-site 568 /note= "Potential glycosylation site"
 FT Modified-site 599 /note= "Potential glycosylation site"
 FT Modified-site 604 /note= "Potential phosphorylation site"
 FT Modified-site 625 /note= "Potential phosphorylation site"
 FT Modified-site 633 /note= "Potential phosphorylation site"
 FT Modified-site 634 /note= "Potential phosphorylation site"
 FT Modified-site 648..733 /note= "Potential phosphorylation site"
 FT Domain /label= PDZ domain signature
 FT Modified-site 690 /note= "Potential phosphorylation site"
 FT Modified-site 705 /note= "Potential phosphorylation site"
 FT Modified-site 721 /note= "Potential phosphorylation site"
 FT Domain /note= "Potential glycosylation site"
 FT /label= PDZ domain signature
 FT Region 819..821 /label= Cell attachment sequence
 FT Modified-site 835 /note= "Potential phosphorylation site"
 XX WO200034477-A2.
 PX 15-JUN-2000.
 PD 10-DEC-1999; 99WO-US30408.
 XX 11-DEC-1998; 98US-0210083.
 PR 09-FEB-1999; 98US-9123456.
 PR 16-MAR-1999; 99US-0119365.
 XX (INCY-) INCYTE PHARM INC.
 XX Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J;
 PI Lu DAM, Azimzai Y;
 XX WPI: 2000-423423/36.
 DR N-PSDB; AAA47424.

XX New human neuron-associated proteins and polynucleotides encoding them,
 PT useful for diagnosis, treatment and prevention of cell proliferative
 PT disorders including cancer, neuronal and neurological disorders
 XX Claim 1; Page 105-108; 145pp; English.
 PS Human neuron-associated proteins (NEUPAP) can be used for for
 CC treating or preventing a disorder associated with decreased
 CC expression or activity of NEUPAP. Antagonists of NEUPAP are useful for
 CC treating or preventing disorder associated with increased expression
 CC or activity of NEUPAP. NEUPAP or their fragments or derivatives are
 CC useful for treating neurological disorder such as epilepsy, ischemic
 CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
 CC disease, Pick's disease, Huntington's disease, dementia and
 CC Parkinson's disease. NEUPAPs are also useful for treating other
 CC demyelinating diseases, bacterial and viral meningitis, prion
 CC diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
 CC metabolic diseases of the nervous system, neurofibromatosis, other
 CC developmental disorders of the central nervous system, cerebral
 CC palsy, neuroskeletal disorders, autonomic nervous system disorders,
 CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and
 CC other neuromuscular disorders, peripheral nervous system disorders,
 CC inherited, metabolic, endocrine, and toxic myopathies, mental
 CC disorders including mood, anxiety and schizophrenic disorders, a cell
 CC proliferative disorder such as actinic keratosis, arteriosclerosis,

CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
 CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
 CC haemoglobinuria, cancers of the adrenal gland, bladder, bone,
 CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
 CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
 CC disease, adult respiratory distress syndrome, allergies, ankylosing
 CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
 CC complications of cancer, hemodialysis, and extracorporeal circulation,
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic
 CC infections, and trauma. This protein was given the Incyte ID no.
 CC 3201753CD1.

XX Sequence 856 AA;

SQ Alignment Scores:

Pred. No.: 9.63e-14 Length: 856
 Score: 263.00 Matches: 146
 Percent Similarity: 36.47% Conservative: 75
 Best Local Similarity: 24.09% Mismatches: 178
 Query Match: 6.59% Indels: 207
 DB: 21 Gaps: 26

US-09-502-945-5 (1-2162) x AAB01383 (1-856)

QY 11 TCGCGTGGCGGCTCTTCCAGCTCTCGGAGCGCGGACCCGAA-----GGAACG 61
 |||:||||| :||| ||||| ||||| |||||
 Db 371 SerGlnSerLeuMetValProGlnSerGlySerProGluProGluSerIleArgAsnThr 390
 QY 62 GGTGCTGCAAGCAGCGAGCTGGACCTGCCAGCCATGGACGAAAGTGGCCGAGAA 121
 |||:||||| :||| ||||| ||||| |||||
 Db 391 SerArgSerSerThrProAlaIlePheAla-Ser--ApproAlaThrCysProIleI 409
 QY 122 TCCGCGCATAGGTGGATTTCTGATTGAAATGATGC----- 158
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 409 eProGly-----CysGluThrThrIleGluIleSerLy 420
 QY 159 ----AGAGAGGACTACTCTATGATGCTG----- 186
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 420 sGlyArgThrGlyLeuGlyLeu-SerIleValGlyGlySerAspThrLeuLeuGlyAlaI 440
 QY 187 -----CGAATGTATCCACCAG-----ACCATGGAGTGGCGCTGCTCGTGG 226
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 440 leIleIleHisGluValTyroGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaG 460
 QY 227 GAGACCTGAAGCTGGTGCATCAAT----- 249
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 460 lyAspGlnIleGluGluValAsnGlyIleAspLeuArgLysAlaThrHisAspGluAlaI 480
 QY 250 -----GAACCCAGCGCTGCTCTGCTCTGTTGTGATGCCATTCGGCGCTGA 292
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 480 leAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyArgAspGluA 499
 QY 293 TCCCACTGAAGCAGCAGGTGGAATATGATCAGCTGACCCCGCGCTCCAGGAAGCTGA 352
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 499 laProTyTyLysGluGluGluValCysAspThrLeuThr----- 511
 QY 353 AGGAGGTGGCTGCTGACCGCTCTGCACCCGAGCGCTCGGCTGAGTGTGGTGGCGCC 412
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 512 ----IleGluLeuGlnLysLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA 530
 QY 413 TGGAGTTGGCTGCGGCTCTTCATCTCCACCTCATCAAGAGCGGTGAGCAGACGCG 472
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 530 rgAsn--AspThrGlyValPheValSerAspIleValLysGlyGlyIleAlaAspAlaA 549
 QY 473 TCGGG---CTCCAGGTAGGCGAGGATCGCCGATCAATGATATTCATCTCCCTCCT 529
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 549 spGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGlyGluAspValArgAsnA 569
 QY 530 GTACCCATGAGGAGGTGATCAACCTCATTCGA---ACCAAGAAAACGTGTGCTCATCAAG 586
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 569 laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGlyThrValThrLeuGluV 589
 QY 587 TGAGACATCGGCTGATCCCGCTGAAAGACTCT-----CCTGATGAG----- 630

Db 842 IleLeuLysarg 845

Search completed: March 21, 2003, 12:44:18
Job time : 73.8072 secs

Db 589 alGlyArgIleLysAlaGlyPropheHisSerGluArgArgProSerGlnSerSerGlnv 609
QY 631 -----
Db 609 alSerGluGlySerLeuSerSerPheThrPheProLeuSerGlySerSerThrSerGluS 629
QY 659 TTGTGTCGGAATCTGGGGCGTCGAGGACGCTGGGGTCC-----CCTGGAATCGGG 712
Db 629 erLeu--GluSerSerLysLysAsnAlaLeuAlaSerGluIleGlnGlyLeuArgT 648
QY 713 AAAACAAGAGAGAAGGCTTTCATCAGCCTGGTAGGCTCCCGAGGCGCTGGCTGCAGCA 772
Db 648 hrValGluMetLysLys-----GlyProThrAspSerLeuGlyIleSerI 663
QY 773 TTTCAGCGGC-----CCATCCAGAAGCCCTGGCATCTTTATCAGCCATGTGAAC 823
Db 663 leAlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetMetHisP 683
QY 824 CTGGCTCCCTCTGCTGCTGAGTG---GGATTGGAGATAGGGGACGACAGATTCTCGAAGTCA 880
Db 683 roThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleC 703
QY 881 ATGGCTCGACTTCTTAACCTGATCACAGAGGCTGTAAATCTGCTGAAAAATAGCC 940
Db 703 ysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValAsnLeuLeuLysAsnAla- 722
QY 941 GCAGCCTGACCATCTCCATTGTAGCTGAGCTGGCCGGGAGCTGTTTCATGACAGACCGGG 1000
Db 723 --SerGlySerIleGluMetGlnValAlaGly-GlyAspValSerValValThrGly 741
QY 1001 AGCGCTGGCAGAGCGGGCAGCTGAGCTGCAGCGGAGGCTTCTCATGCAGAGC 1060
Db 741 ----- 741
QY 1061 GGCTGGCGATGGATCCCAAGAATCTCTCAGAGCAGCAGGAGATGGAGCGCAAGGA 1120
Db 741 ----- 741
QY 1121 GAAAGAAATTGCCAGAGCAGCAGAGGAAATGAGAGATACCGGAGGAGATGGAAC 1180
Db 741 ----- 741
QY 1181 AGATTGTAGAGGAGAGAGAAGATTAAAGAACAATGGAGAAGACTGGGGCTCAAAGG 1240
Db 741 ----- 741
QY 1241 AACAGCTACTCTGCTTAACCATCATCTGCTGAGGTACACCCAGTACCCCTTCGCAAGC 1300
Db 742 -----HisGlnGlnGluProAlaSerSer 749
QY 1301 CAAAGTGATTCCGGAATATGAGNAGGCTTGACCCCTACTCTATGTTCACC----- 1354
Db 750 SerLeuSerPhe-----ThrGlyLeuThrSerSerSerIlePheGlnAspAsp 765
QY 1355 -----CCAGACGACATCATGGGAAGGATGTCGGCTCCCTACGCAATCAAGAGGAGGA 1408
Db 766 LeuGlyProProGlnCys-----LysSerIleThrLeuGluArg-----GlyProAsp 781
QY 1409 TCCTTAGACCTGGCCGGAAGGCGGTGGACTCCCGCATTTGG-----AAGGTGGTC 1462
Db 782 GlyLeuGlyPheSerIleValGlyTyrglySerProHisGlyAspLeuProIleYr 801
QY 1463 GTTCTGTGTATGAGCGGGAGCTGCTGAGCGGCGATGGTGGCATTTGTAAGGGGAC 1522
Db 802 VallysThrValPheAlaValAlaAlaSerGluAspGlyArgLeuLysArgGlyAsp 821
QY 1523 GAGATCATGGCAATCAACCGCAAGATTGTGACAGACTACACCCCTGGCTGAGGCTGACGCT 1582
Db 822 GlnIleIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaValAla 841
QY 1583 GCCCTGCAGAG 1594
| | | | |

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run On: March 21, 2003, 12:34:32 ; Search time 16.7489 Seconds
(without alignments)
7595.986 Million cell updates/sec

Title: US-09-502-945-5
Perfect score: 3989
Sequence: 1 cctggccggtcggtgcg.....tattttccagcttaaaaaa 2162

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer_n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09502945/runat_14032003_101059_19145/app_query.fasta_1.10979
-DB-Issued Patents AA -QFWT=fastan -SURFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN1_1_80.#runat_14032003_101059_19145 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*\n2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*\n3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*\n4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*\n5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*\n6: /cgn2_6/ptodata/1/1aa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	263	6.6	2037	4	US-09-306-998-3
2	250.5	6.3	1881	4	US-09-233-086-3
3	229.5	5.8	724	4	US-09-562-737-21
4	211	5.3	724	4	US-09-562-737-28
5	210.5	5.3	724	4	US-09-562-737-23
6	206.5	5.2	1050	3	US-09-045-632-50
7	204.5	5.1	2466	3	US-09-080-855-12
8	204.5	5.1	2466	5	PCT-US94-09943-2
9	204.5	5.1	2485	4	US-09-290-640-46
10	203.5	5.1	1050	3	US-09-045-632-49
11	202	5.1	1112	3	US-09-045-632-2
12	202	5.1	1112	3	US-09-045-632-3

13	201	5.0	604	3	US-09-045-632-14	Sequence 14, Appl
14	201	5.0	702	3	US-09-045-632-15	Sequence 15, Appl
15	201	5.0	1018	3	US-09-045-632-16	Sequence 16, Appl
16	201	5.0	1061	3	US-09-045-632-32	Sequence 32, Appl
17	199	5.0	724	4	US-09-562-737-22	Sequence 22, Appl
18	199	5.0	2465	2	US-08-596-291-3	Sequence 3, Appl
19	199	5.0	2465	3	US-09-100-804-3	Sequence 3, Appl
20	194	4.9	724	4	US-09-562-737-30	Sequence 30, Appl
21	193.5	4.9	450	2	US-08-665-037-2	Sequence 2, Appl
22	193.5	4.9	450	2	US-08-666-067-2	Sequence 2, Appl
23	193.5	4.9	450	2	US-08-732-870-2	Sequence 2, Appl
24	190	4.8	724	4	US-09-562-737-26	Sequence 26, Appl
25	188.5	4.7	724	4	US-09-562-737-24	Sequence 24, Appl
26	186.5	4.7	724	4	US-09-562-737-29	Sequence 29, Appl
27	186	4.7	507	3	US-09-045-632-13	Sequence 13, Appl
28	181	4.5	724	4	US-09-562-737-27	Sequence 27, Appl
29	179.5	4.5	519	3	US-08-997-445D-2	Sequence 2, Appl
30	172.5	4.3	960	4	US-09-219-849-6	Sequence 6, Appl
31	172	4.3	724	4	US-09-562-737-25	Sequence 25, Appl
32	171.5	4.3	502	3	US-09-045-632-24	Sequence 24, Appl
33	171.5	4.3	602	3	US-09-045-632-20	Sequence 20, Appl
34	171.5	4.3	818	3	US-09-045-632-25	Sequence 25, Appl
35	171.5	4.3	861	3	US-09-045-632-34	Sequence 34, Appl
36	171.5	4.3	918	3	US-09-045-632-21	Sequence 21, Appl
37	171.5	4.3	961	3	US-09-045-632-33	Sequence 33, Appl
c	171	4.4	552	4	US-09-219-849-7	Sequence 7, Appl
	169	4.2	552	4	US-09-219-849-7	Sequence 7, Appl
40	166	4.2	283	3	US-09-045-632-27	Sequence 27, Appl
41	166	4.2	599	3	US-09-045-632-28	Sequence 28, Appl
42	166	4.2	642	3	US-09-045-632-35	Sequence 35, Appl
43	161.5	4.0	610	1	US-08-410-804-1	Sequence 1, Appl
44	161.5	4.0	610	1	US-08-259-514-1	Sequence 1, Appl
45	161.5	4.0	610	2	US-08-858-311-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-306-998-3
; Sequence 3, Application US/09306998
; Patent No. 6291173
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavligian, Sean V.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/306,998
; CURRENT FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 60/084,740
; EARLIER FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-306-998-3

Alignment Scores:
Pred. No.: 3e-14 Length: 2037
Score: 263.00 Matches: 146
Percent Similarity: 36.47% Conservative: 75
Best Local Similarity: 24.09% Mismatches: 178
Query Match: 6.59% Indels: 207
DB: 4 Gaps: 26

US-09-502-945-5 (1-2162) x US-09-306-998-3 (1-2037)

Qy 11 TCGGGTTCGGCGCTTTTCACGCTCTCGACGCGGCGACCCGAA-----GGAAACG 61
|||||:|||||::: ||| ||||| |||||
Db 1552 SerGlnSerLeuMetValProGlnSerGlySerProGluProGluSerIleArgAsnThr 1571
|||
Qy 62 GGTGTCGACGACGACGCTGGACCTGGCCCGACCATGGACCGGAAAGTGGCCCGAGAA 121

Db 1572 SerArgSerSerThrProAlaIlePheAla-Ser---AspProAlaThrCysProIleIl 1590
QY 122 TCCGCGCATAAAGTGGATTTCCTGATTGAAATGATGC 158
Db 1590 eProgly-----CysGluThrThrIleGluIleSerLy 1601
QY 159 -----AGAGAAGGACTATCTCTATGATGTGCTG- 186
Db 1601 sGlyArgThrGlyLeuGlyLeu-SerIleValGlyGlySerAspThrLeuLeuGlyAlaI 1621
QY 187 -----CGAATGTACACCCAG-----ACCATGGACCTGCCCGCTCGTGG 226
Db 1621 leIleIleHisGluValTyGluGluGlyAlaAlaCysLysAspGlyArgLeuThrPalag 1641
QY 227 GAGACCTGAAGCTGTGCTCAAT- 249
Db 1641 lyAspGlnIleLeuGluValAsnGlyIleAspLeuArgLysAlaThrHisaspGluAlaI 1661
QY 250 -----GAACCCAGCCGCTGCTCCTCTGTTTGTATGCCATTCGGCGCTGA 292
Db 1661 leAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyArgAspGluA 1680
QY 293 TCCCACTGAAGCACCAGGTGGAAATGATGATCAGCTGACCCCGCGGCTCCAGGAAGCTGA 352
Db 1680 laProTyrLysGluGluGluValCysAspThrLeuThr 1692
QY 353 AGGAGTGTCTGGACCTGCTGCACCCGAGGCTCGCTGAGTGTGCTGTGGCC 412
Db 1693 -----IleGluLeuGlnLysLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA 1711
QY 413 TGGAGTTTGGCTGTGGCTCTTCATCTCCACCTCATCAAAAGCGGTTCAGGACAGCGG 472
Db 1711 rGAsn---AspThrGlyValPheValSerAspIleValLysGlyGlyIleAlaAspAlaA 1730
QY 473 TCGGG---CTCAGGTAGGGAGGATGTCGGGATCAATGGATATTCCTCTCTCT 529
Db 1730 spGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGlyGluAspValArgAsnA 1750
QY 530 GTACCATGAGGAGTCATCAACCTCATTCGA---ACCAAGAAACTGTGTCATCAAG 586
Db 1750 laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGlyThrValThrLeuGluV 1770
QY 587 TGAGACATATCGGCTGTATCCCGTGAAGTCT- 630
Db 1770 alGlyArgIleLysAlaGlyProPheHisSerGluArgProSerGlnSerSerGlnV 1790
QY 631 -----CCCTCACTTGGCAGTATGTGGATCACT 658
Db 1790 alSerGluGlySerLeuSerSerPheThrPheProLeuSerGlySerSerThrSerGluS 1810
QY 659 TTGTGTGCGGAATCTGGGGCGTCCGAGGAGCCTGGCTCC-----CCTGGAAATCGGG 712
Db 1810 erLeu---GluSerSerSerLysLysAsnAlaLeuAlaSerGluIleGlnGlyLeuArgT 1829
QY 713 AAAACAAGGAGAAGAAGTCTTCATCAGCCTGTGAGTCCCGAGGCTTGGCTGCGAGCA 772
Db 1829 hrValGluMetLysLys-----GlyProThrAspSerLeuGlyIleSerI 1844
QY 773 TTTCCAGCGGC-----CCCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAAC 823
Db 1844 leAlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetMetHisP 1864
QY 824 CTGGCTCCCTGTCTGTAGGTG---GGATTGGAGATAGGGACCATGTGCGAAGTCA 880
Db 1864 roThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleC 1884
QY 881 ATGGCTCGCACTCTCTAACTGGATCACAAGGAGGCTGTAATGTCTCTCAAAATAGCC 940
Db 1884 ysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValAsnLeuLeuLysAsnAla- 1903
QY 941 GCAGCTGACCATCTCCATTTAGTGTACGTGACCTGCGCGGAGCTGTTCATGACAGCCGG 1000

Db 1904 --SerGlySerIleGluMetGlnValValAlaGly-GlyAspValSerValThrGly 1922
QY 1001 AGCGCTGGCAGAGCGCGCGAGCGTGCAGCTGCAGCGGAGGAGCTTCTCATCAGAAGC 1060
Db 1922 ----- 1922
QY 1061 GGCTGGCGATGGAGTCCAAACAAGATCTCTCCAGGAGCAGCAGGATGGAGCGGCAAGGA 1120
Db 1922 ----- 1922
QY 1121 GAAAGAAATTTCCCGAAGGACGACAGAGGAAATGAGAGATACCGGAAGGATGGAAC 1180
Db 1922 ----- 1922
QY 1181 AGATTGTAGAGGAGGAGAGAAGTTTAAAGCAATGGGAAGACTGGGGCTCAAAGG 1240
Db 1922 ----- 1922
QY 1241 AACAGCTACTTTCGCTTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGC 1300
Db 1923 -----HisGlnGlnGluProAlaSerSer 1930
QY 1301 CAAAGTGATTTCGGAATATATGAGAGGGCTTTGACCCCTACTCTATGTTCCACC----- 1354
Db 1931 SerLeuSerPhe-----ThrGlyLeuThrSerSerSerIlePheGlnAspAsp 1946
QY 1355 -----CCAGAGCAGATCATCGGGAAGGATGCTCGGCTCTCCATCAGCATCAAGAAGAGGGA 1408
Db 1947 LeuGlyProGlnCys-----LysSerIleThrLeuGluArg-----GlyProAsp 1962
QY 1409 TCCTTAGACCTGGCCCTGGAAGGGGTGAGTGCATCCCTCCATCAGCATCAAGAAGAGGGA 1462
Db 1963 GlyLeuGlyPheSerIleValGlyTyGlySerProHisGlyAspLeuProIleTyr 1982
QY 1463 GTTTCGTGTGTATGAGCGGGAGCTGCTGAGCGGCATGGTGGCATTTGAAAGGGGAC 1522
Db 1983 ValLysThrValPheAlaLysGlyAlaAlaSerGluAspGlyArgLeuLysArgGlyAsp 2002
QY 1523 GAGATTCATGCAATCAACGCAAGATTTGACAGACTACACCTGGCTGGCTGAGCTGACGCT 1582
Db 2003 GlnIleIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaVala 2022
QY 1583 GCCCTGCGAGAAG 1594
Db 2023 IleLeuLysArg 2026
RESULT 2
US-09-233-086-3
; Sequence 3, Application US/09233086
; Patent No. 6337192
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigan, Sean V.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: MMSCl - An MMACl Interacting Protein
; FILE REFERENCE: MMSCl Gene
; CURRENT APPLICATION NUMBER: US/09/233,086
; CURRENT FILING DATE: 1999-01-19
; EARLIER APPLICATION NUMBER: US 60/071,861
; EARLIER FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-233-086-3
Alignment Scores:
Pred. No.: 3,56e-13 Length: 1881
Score: 250.50 Matches: 155
Percent Similarity: 36.03% Conservative: 117
Best Local Similarity: 20.53% Mismatches: 281

Db	1503	agInGlyLeuValGlnLeuGluIleclYargLeuArgAlaGlySer	---trpThr----	1620
QY	1915	AGCAGCCAGCCGCTTTGGGAGAAGATCTCAAGGATCCAGACTCTCATTTCTTCTCTCTGGC	1974	
Db	1621	SerAlaArgThrThrSerGlnAsnSerGlnGlySerGlnSerAlaHisSerSer	---1639	
QY	1975	CCAGTGAATTTGGTCTCTCCACGCTTTGGGGACTCCTTT	2013	
Db	1640	-----CvshHisProserPheAlaProValIle	1648	

RESULT 3

```

US-09-562-737-21
; Sequence 21, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-21

```

Alignment Scores:

Fragment Scores:	
Pragm. No.:	1.42e-11
Score:	229.50
Length:	724
Matches:	146
Conservative:	101
Mismatches:	204
Indels:	207
Gaps:	35
DB:	4
US-09-502-945-5 (1-2162) x US-09-562-737-21 (1-724)	

Qy	340	TCCAGGAAGCTGAAGAGGTCGCTCGAGCCGTCGCACCCGAAAGCGCTGGCGCTGAGT	399
Db	155	AlaGluLysIleLeuGluIleLysLeuIleLys---GlyProLysLeuGlyLysPheSer	173
Qy	400	GTGCGTGGTGGC-----CTGGAGTTTGGCTGTGGGCTCTTCATCTCCAC	444
Db	174	IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys	193
Qy	445	CTCATCAAGAGCGGTGAGGACAGACGGCTGGG---CTCAGGTAGGGGACGAGATCGTC	501
Db	194	IleIleGluGlyGlyAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu	213
Qy	502	CGGATCAATGATGATTCCATCTCTCTCTGTACCCATGAGGAGTCATCAACCTCATTCGA	561
Db	214	AlaValAsnSerValGlyLeuGluAspValMetHisLysAlaValAlaLeuLys	233
Qy	562	---ACCAAGAAAACGTGTGTCATCAAGTG-----AGACATCGCGCTGATCCCC	609
Db	234	AsnThrTyrAspValValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp	253
Qy	610	GTGAAAGCTCTCGTCGATGAGCCCTCAGTTG---CAGTATGTGATCAGTTTGTGTGG	666
Db	254	SerTyrAlaProProAspIleThrThrSerTyrSerGlnHisLeuAspAsnGluIleSer	273
Qy	667	GAATCTGGGGCGTGCAGGCGAGCTGGGC-----	696
Db	274	HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr	289
Qy	697	-----TCCCTGGAAATCGGAA-----AACCAAG	720

QY 1636 CGCTGCCCCCAAGGAGTATGACGATGAGCTGACCTTCTTGTGCAAGTCCAAAGGGG 1695
Db 592 -----AspIleGlnAlaHisLysPheIleGluAlaGlnTy 604
QY 1696 AAACCAAAATTCACGGTTAGGAACAGT-----GAGCTCGG 1731
Db 604 rAsnSer---HisLeuTyrGlyThrSerValGlnSerValArgGluValAlaGluGlnG1 623
QY 1732 GCGCCACCTCGTGAACAAAGCCCTCGGACCGCTTGAGAGAGGCCACATGACACACAC 1791
Db 623 yIshisCysIleLeuAspValSerAlaAsnAlaValArgGluGlnAlaAlaHisLe 643
QY 1792 CAGATGGCATCTGGGACCTGAATCTATCACCAGGAATCTCAACTCCCTTTGGCCCT 1851
Db 643 u-----HisProIleAlaIlePheIleArgProArgSerLeuGluAsnValLeuGluI1 661
QY 1852 GAAC-----CAGGCCAGATAAGGAACAGCTCGGGCCACTTTTGA 1893
Db 661 eAsnLysArgIleThrGluGluGln-AlaArgLysAlaPheAspArgAlaThrLysLeuG 681
QY 1894 AG-----GCCAATGTGGAGGAAGGAGGAGCAGCCAGCCGTTT 1929
Db 681 LuGlnGluPheThrGluCysPheSerAlaIleValGluGlyAspSerPheGluGluIleT 701
QY 1930 GCGAGAGATCTCAGG---ATCCAGACTCTCATTT---CCTTCTCTCTGG 1973
Db 701 yHisLysValLysArgValIleGluAspLeuSerGlyProTyrIleTrp 717

RESULT 4

US-09-562-737-28
; Sequence 28, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UFSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-03-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-28

Alignment Scores:
Pred. No.: 5,91e-10 Length: 724
Score: 211.00 Matches: 139
Percent Similarity: 39.74% Conservative: 101
Best Local Similarity: 23.01% Mismatches: 205
Query Match: 5.29% Indels: 159
Db: 4 Gaps: 35

US-09-502-945-5 (1-2162) x US-09-562-737-28 (1-724)

QY 250 GAACCCAGCGCTGCTGCTCTGTTGATGCCATTCGG-----CCGCTGATC 294
Db 107 GlnGlnGlyArgLeuArgValAsnAspSerIleArgPheValAsnGluValAspValArg 126
QY 295 CCACTGAACGACACAGGTGGAATATGATCAGCTGACCCCG-----333
Db 127 GluSerThrHisSerAlaValGluAlaLeuThrGluAlaGlySerIleValArgLeu 146
QY 334 -----CGGCGC-----TCCAGGAAGCTGAAGAGGTGCGTCTGGACCGCTG 375
Db 147 TyrTrpMetArgArgLysProProAlaGluLysTyrIleGluIleLysLeuIleLys--- 165

QY 376 CACCCCGAAGCCCTCGGCTGAGTGTGGTGGTGGC-----CTGGAGTTT 420
Db 166 GlyProAlaGlyLeuGlyPheSerIleAlaGlyAspGlyAsnGlnHisIleProGly 185
QY 421 GGCTGTGGCTCTCATCTCCCATCATCAAGCGGCTAGGCAGACAGCGTCCGG--- 477
Db 186 AspAsnGluIleTyrValThrLysIleIleGluGlyPheAlaAlaHisLysAspGlyArg 205
QY 478 CTCCAGGTAGGGGACGAGATCGTCCGGATCAATGATATTCATCTCTCTCTGACCCAT 537
Db 206 LeuGlnGlyGlyAspLysIleLeuAlaValAsnSerHisGlyLeuGluAspValMetHis 225
QY 538 GAGAGGTCACTCAACCTCATTCGA---ACCAGAAGAACTGTCTCCATCAAAAGTGAGACAC 594
Db 226 GluAspIleValAlaAlaLeuLysAsnThrTyrAspLysValTyrLeuLysValAlaLys 245
QY 595 ATCGGCTCTCCCTCGAAAGCTCT-----CTGATGAGCCCTCCTCCT---TGG 642
Db 246 ProSerLeuAlaTyrLeuSerAspSerTyrAlaProMetAspIleThrThrSerTyrSer 265
QY 643 CAGTATGTGGATCATGTCTCGGAATCTGGGGCGGTGCGAGCAGC----- 690
Db 266 GlnHisAsnAspAsnGluIleSerHisSer---SerTyrGlnGlyThrAspTyrProThr 284
QY 691 -----CTGGCTCCCTCGAAATCGGAA----- 714
Db 285 AlaMetThrArgThrSerProArgArgTyrSerProValSerLysAspLeuLeuGlyGlu 304
QY 715 -----AACAGGAGGAAGAGTCTTCATCAGCTGTGAGGTCCCGAGTCTGGC 765
Db 305 GluAspIleThrArgGluProArgArgIleValIleHisValGlySerThrGlyLeuGly 324
QY 766 TGCAGCATTTCCAGGGGCCCATCCAGAGCTCGCATCTTTTATCAGCATGTG--- 819
Db 325 PheAsnIleTrpGlyGlyGlu---AspGlyGlyIlePheTyrSerPheIleLeuAla 343
QY 820 ---RAACCTGCTCCCTGCTCTGCTGAGGTGGATTGGAGATAGGGGACCATGTGCGAA 876
Db 344 GlyGlyProAlaAlaLeuSerGlyGlu-----LeuArgLysGlyAspAspIleLeuSer 361
QY 877 GTCATATGCGCTCGACTTCTCTTAACCTGGATCACAAGGAGGTGTAATGTCTGCTGAAAAT 936
Db 362 ValAsnGlyValAspLeuGluAsnAlaSerHisGluGlnAlaAlaIlePheLeuLysAsn 381
QY 937 AGCCGACCTTGACCATCTCCATTTGATGCTGCGAGTGGCGGAGCTGTTCTATGACACAC 996
Db 382 Ala---GlyGlnThrValThrGlyIleAlaGln-----TyrLysProGlu 395
QY 997 CGGAGCGGCTGGCAGAGGCGCGCAGCTGAGCTGCGAGCGGAGGAGCTTCTCATGACAG 1056
Db 396 GluTyrHisArgPheGluAlaLysIleHisAspLeuIleGluGln-----LeuMetAsn 413
QY 1057 AAGCGGCTG-----GCGATGGAGTCCAAACAGATCCTCCAG----- 1092
Db 414 SerSerLeuGlyLysGlyThrAlaSerLeuArgSerAsnProLeuArgGlyPheTyrIle 433
QY 1093 -----GAGCAGCAGAGATGGAG 1110
Db 434 ArgAlaLeuPheMetTyrAspLysThrLysAspCysGlyPheAsnSerGlnAlaLeuSer 453
QY 1111 CGGCAAGAGAGAAAGAAATTCGCCAGAGGACAGGAGGAGAAATGAGAGATACCGG--- 1167
Db 454 PheHisPheGlyGlnValLeuHisValIleAspAlaSerAspArgGluTrpIleAla 473
QY 1168 -----AAGGAGATGGAACAGATT-----GTAGAGGAGAGAGAGAG 1203
Db 474 ArgArgValHisThrAspSerGluThrAspAspIleGlyPheValProSerLysArgArg 493
QY 1204 TTATAGACCAATGG-----GAAGAAGACTGG----- 1230
Db 494 ValGluArgArgTrpSerArgLeuLysAlaLysAspTrpTyrSerSerSerGlySer 513
QY 1231 ---GGCTCAAGGAACAGCTACTCTTGGCTTAAACCATCACT---GCTGAGGTACAC--- 1281


```
QY 1237 -----AAGGAACAGCTACTTGTCTGCTAAACCATCACT---GCTGAGTA 1278
Db 512 GlyGluGlnGlyArgGluAspSerValLeuSerPheGluThrValThrGlnMetGluVal 531
QY 1279 CACCCAGTACCCTTCGCAAGCCAAAGTGATTTCGCGAAATATGAGGAAGGCTTTGAGCC 1338
Db 532 HisGlyAlaArg-----Pro 536
QY 1339 CTACTCTATGTTACCCAGCAGCAGATCATGCGGAAGGATGTCGGGTCTCCTACGCATCAA 1398
Db 537 IleIleIleLeuGlyProHisLysAspArgAla----- 547
QY 1399 GAAGGAGGATCCTTAGACCTGGCCCTGGAAGGGGTGTGGACTCCGCCATTTGGGAAGGT 1458
Db 548 AsnAspAspLeuLeuIleGluPheProAspLys----- 558
QY 1459 GGTGCTTTCTGCTGTATGACGGGGAGCTGCTGAGCGGCATGTTGGCATTGTGAAGG 1518
Db 559 -----PheGlySerCys-LysProHisThrThrArgProLysAr 571
QY 1519 GGACGAGATCATGCAATCAACGCCAAGATTGTGACAGACTACACCTGGCTGAGGCTGA 1578
Db 571 gGlu---LeuGluIleAspGlyArg-----AspTyrHisPheMetSerSerAr 586
QY 1579 CGTGCCTCGCAGAGGCGCTGGAATCAGGCGGGGACTGGATCGACCTTGTGGTTGCCGT 1638
Db 586 gGluLysMetGluLys----- 591
QY 1639 CTGCCCCCACAAGGATGATGACGATGAGCTGACCTTCTGCTGAAGTCCAAAAGGGGAAA 1698
Db 592 -----AspAsnGlnAlaHisLysPheIleGluAlaGlyArgTyrAs 605
QY 1699 CCAATTCACCGCTTAGAAACAGT-----GAGCTCCGGGCC 1734
Db 605 nSer---HisLeuTyrGlyThrSerSerGlnSerValArgGluValAlaGluGlnThrLy 624
QY 1735 CCACCTCTGTGAACAAAGCCTCGGACCGCTTGAGAGCGGCACATGACACACACAGG 1794
Db 624 SHiscysIleLeuAspValSerAlaValAlaValArgLeuGlnAlaHis----- 642
QY 1795 ATGCATCTTTGGGACCTGAATCATCACCCAGGAATCTCAAACTCCCTTTGGCCCTGAA 1854
Db 643 -TriHisProIleAlaIlePheIleArgProTyrSerLeuGluAsnValLeuGluIleAs 662
QY 1855 C-----CAGGCCAGATAAGAACAGCTCGGGCCACTTTTGTGAAG- 1895
Db 662 nAlaArgIleThrGluGluGln-AlaArgLysAspPheAspArgAlaThrLysLeuGluG 682
QY 1896 -----GCCAATGTGGAGGAAGGAGCAGCCAGCCCTTTGGG 1932
Db 682 InPhePheThrGluCysPheSerAlaIleValGlyLysAspSerPheGluIleIleYrH 702
QY 1933 AGAAGATCTCAAGATC-----CAGACTCTCATCTCTCTCTCTCTG 1973
Db 702 IsHisValLysArgValIleGluAspLeuSerIleProTyrIleTyr 717
```

RESULT 6

```
US-09-045-632-50
; Sequence 50, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Haganir, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
```

```
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/045,632
FILING DATE: 19-MAR-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/041,016
FILING DATE: 19-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Corless, Peter F.
REGISTRATION NUMBER: 33,860
REFERENCE/DOCKET NUMBER: 48147/1699-CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1050 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-045-632-50
```

```
Alignment Scores:
Pred. No.: 1,81e-09 Length: 1050
Score: 206.50 Matches: 189
Percent Similarity: 35.22% Conservative: 99
Best Local Similarity: 22.21% Mismatches: 229
Query Match: 5.18% Indels: 229
DB: 3 Gaps: 35
```

```
US-09-502-945-5 (1-2162) x US-09-045-632-50 (1-1050)
QY 322 CAGCTGACCCCGCGCTCCAGGAAGCTGAAGAGGTGCTGTCGACCGCTGTCACCCC 381
Db 77 GlnSerIleProGluGluPheArgGlyIleThrMetValGluLeuIleLysArgGluGly 96
QY 382 GAAGGCTCCGGCTGAGTGTGCTGGTGGCTGGAGTTGGCTGGCTTCATCTCC 441
Db 97 SerThrLeuGlyLeuThrIleSerGlyGlyThrAspLysAspGlyLysProArgValSer 116
QY 442 CACCTCATCAAAAGCGGT---CAGGCAGACAGCGCTCGGGCTCCAGGTAGGGGACGAGATC 498
Db 117 AsnLeuArgProGlyGlyLeuAlaAlaArgSerAspLeuLeuAsnValGlyAspTyrIle 136
QY 499 GTCCGGATCAATGATATTCATCTCCTCTCTACCCATGAGGAGGTATCAACCTCAT 558
Db 137 ArgSerValAsnGlyIleArgLeuThrArgLeuArgHisAspGluIleIleThrLeuLeu 156
QY 559 CGAAC----- 564
Db 157 LysAsnValGlyGluArgValValLeuGluValGluTyrGluLeuProProAlaPro 176
QY 565 -----AAGAAACGTGCTCCATCAAAAGT-----AGA 591
Db 177 GluAsnAsnProArgIleIleSerLysThrValAspValSerLeuTyrLysGluGlyAsn 196
QY 592 CACATCGGCTGATCCCGCTGAAAGCTCTCCTGATGAG-----CCCTC 636
Db 197 SerPheGlyPheValLeuArgGlyGlyAlaHisGluAspLeuHisLysSerArgProLeu 216
QY 637 ACTGGCAGTATGTG----- 651
Db 217 ValLeuThrTyrValArgProGlyGlyProAlaAsnArgGluGlySerLeuLysValGly 236
QY 652 GATCAGTTTGTGCGGAATCTGG----- 675
```



```

Percent Similarity: 36.35%      Conservative: 94
Best Local Similarity: 21.93%   Mismatches: 214
Query Match: 5.13%             Indels: 201
DB: 3                           Gaps: 29

US-09-502-945-5 (1-2162) x US-09-080-855-12 (1-2466)

QY 1 CTTGGCCCGCTCGCGTCCGCGCTCTT----- 27
DB 991 ProProGlnThrValAlaGluLeuValGlyLysProSerHisGlnMetSerArgSer 1010
QY 28 ---TCCAGTCTCTGGCAGCCGCGCCACCCGGAAGGAGCGGTCTGTCACACGACGACGTGA 84
DB 1011 AspaIaGluSerLeuAlaGlyValThrLysLeuAsnAsnSerLysSerValAla----- 1028
QY 85 CTTGGCCCGCAGCATGGACCGCAAAAGTGGCCCGAGAAATCCCGCATAAAGTGGATTTCTG 144
DB 1029 -----SerLeuAsnArgSerProGluArgArg---LysHisGluSerAspSerSer 1044
QY 145 ATTGAAATGATGCAGAGAGGACTATCTCTATGATGCTCGGAATGTACACACGACG 204
DB 1045 SerIleGluAspProGlyGlnAlaTyrValLeuAspValLeu-----HisLysArg 1061
QY 205 ATGGAGCTGGCCGCTGCTCTGGGAGACCTGAAGCTGGTCAATCAATGAACCCAGCCGCTCG 264
DB 1062 TrpSer-----IleValSerSerProGluArg--- 1070
QY 265 CTTCTGTTGATGCCATTCCTGGCGGTGATCCACCTGAAGCAGACGAGTGAATATGATCAG 324
DB 1071 -----GluIleThrLeuValAsnLeuLysLysAspAlaLysTyr----- 1083
QY 325 CTGACCCCGCGCTCCAGGAAGCTGAAGAGGTGCTCTGGACCGTCTGACCCCGAA 384
DB 1083 ----- 1083
QY 385 GGCCTCGGCTGAGTGTGCTGCTGGCTGGAGTTTGGC-----TGTGGCTCTTC 435
DB 1084 GlyLeuGlyPheGlnIleIleGlyGlyGlyLysMetGlyArgLeuAspLeuGlyIlePhe 1103
QY 436 ATCTCCACCTCATCAAAAGCGCGTGCAGGACAGCAGCTGGG---CTCCAGGTAGGGAC 492
DB 1104 IleSerSerValAlaProGlyGlyProAlaAspPheHisGlyCysLeuLysProGlyAsp 1123
QY 493 GAGATCGTCCGATCAATGGATATCCATCTCTCTGTACCATCAGGAGTCAAC 552
DB 1124 ArgLeuIleSerValAsnSerValSerLeuGluGlyValSerHisAlaIleGlu 1143
QY 553 CTCATTGCAACCAAGAAACTGTGTCCATCAAGTGAACATCGCGCTGATCCCGTG 612
DB 1144 IleLeu----- 1145
QY 613 AAAAGCTCTCTGATGAGCCCTCACTTGGCAGTATGTGATGAGTGTGCGGAATCT 672
DB 1146 GlnAsnAlaProGluAsp-----ValThrLeuValIleSerGlnPro 1159
QY 673 GGGGCGCTCGGAGGCGCTGGCTCCCTCGAATCGGAAACAAAGAAAGAGGTC 732
DB 1160 -----LysGluLys----- 1162
QY 733 TTTCATCAGCTGTTAGGCTCCGAGCCCTTGCTGCAGCATTTCCAGCGGCCCATCCAG 792
DB 1163 ---IleSer-----LysValProSerThrProValHis 1172
QY 793 AAGCCTGGCATCTTTATCAGCCATGTAAACCTGGCTCTCTGCTGAGTGGGATTG 852
DB 1173 LeuThrAsnGluMetLysAsnTyrMetLysLysSer----- 1185
QY 853 GAGATAGGGGACCATGTTCGAATCAATGCGCTGCGCTCTCTTAACCTGGATCAACAG 912
DB 1186 -----TyrMetGlnAspSerAlaIleAspSerSerLysAspHis 1200
QY 913 GAGGCTGTAATGTCTGAAATAATAGCCGACGCTGACCATCTCCATTTGTAGCTGACCT 972
DB 913 ----- 972

```

1201 TrpSerArgGlyThrLeu-----ArgHisIleSerGluAsnSerPheGlyProSer 1217
 QY 973 GGC-----CGGAGCTGTTTCATGACAGACGCGGAGCGGTGGCAGAGCGCGGACGCT 1026
 DB 1218 GlyGlyLeuArgGluGlySerLeuSerGlnAspSerArgThrGluSer-AlaSerLe 1237
 QY 1027 GAGCTGCAGC-----GGCAGGAGCTTCTCATGTCAGAGCGGC-----TGGCGATG 1071
 DB 1237 userGlnSerGlnValAsnGlyPheAlaSerHisLeuGlyAspGlnThrTrpGlnG 1257
 QY 1072 GAGTCCCAACAGATCCTCCAGGAGCAGCAGGATGGAGCGCAAGAGGAGGAGAAAT 1131
 DB 1257 user-GlnHisGlySerProSerProSerValIleSerLysAlaThrGluLysGluThrP 1277
 QY 1132 GCCCAGAGCAGCAGAGGAAATAGAGATACCGGAA---GGAGATGGACAGATTT- 1187
 DB 1277 heThrAspSerAsnGlnSerLysThrLysProGlyIleSerAspValThrAspTyrS 1297
 QY 1188 -----AGAGGAGGAGAGAAAGTTTAAAGAACCAATGGGAAGAGACTGGGGCTCAAGAA 1242
 DB 1297 erAspArgGlyAspSerAspMetAspGluAla----- 1308
 QY 1243 CAGCTACTCTTCCCTAAACCATCACCTGAGGTACACCCAGTACCCCTTCGCAAGCA 1302
 DB 1308 hrTyrSerSerSerGlnAspHisGln-----ThrProLysGlnLysSerSerS 1325
 QY 1303 AAGTGATTTCCGGAATATAGGAAGGCTTTGACCCCTTACTTATGTTCACCCAGCA 1362
 DB 1325 erValAsnThrSerAsn--LysMetAsnPheLysThrPheSerSerProLysPr 1344
 QY 1363 GATCATGGGAGGATGTCGCGCTCTACGATCAAGAGGAGGAGTCTTAGACCTGGC 1422
 DB 1344 o-----GlyAspIlePheGluValGluLeuAlaLysAsnSerLeuGlyIleSe 1362
 QY 1423 CTGGAAGCGGTGTGGACTCCCCCAT-----GGGAAGCTGTGCTTCTGCTGTGTA 1476
 DB 1362 rValThrGlyGlyValAsnThrSerValArgHisGlyGlyIleTyrValLysAlaVal 1382
 QY 1477 TGAGCGGAGCTGCTGAGCGGCATGTTGGCATGTGAAAGGCGGAGGAGCATGCGCAAT 1536
 DB 1382 eProGlnGlyAlaAlaGluSerAspGlyArgIleHisLysGlyAspArgValLeuAla 1402
 QY 1537 CAACGGCAAGATTTGACAGACTACACCTGCTGAGGCTGAGGCTGAGGCTGCTGCTG 1588
 DB 1402 lAsnGlyValSerLeuGluGlyAlaThrHisLysGlnAlaValGluThrLeuArgAsnTh 1422
 QY 1588 ----- 1588
 DB 1422 rGlyGlnValValHisLeuLeuGluLysGlyGlnSerProThrSerLysGluHisVa 1442
 QY 1589 -----CAGAAGCGCTGGAATCAGGCGGGGAGCTG 1617
 DB 1442 lProValThrProGlnCysThrLeuSerAspGlnAsnAlaGlnGlyGlnGlyProGlu 1462
 QY 1618 GATGACCTTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1668
 DB 1462 sValLysLysThrThrGlnVal-----LysAspTyrSerPheValThrGluGluAs 1479
 QY 1669 GACCTC-----TTGCTGAAGTCCAAAGGGGAAACCAATTCACGCTTAGGAA 1719
 DB 1479 nThrPheGluValLysLeuPheLysAsnSerSerGlyLeuGlyPheSerPheSerArg 1499
 QY 1720 CAGTGAGCTCGCGCCACCTCGTGAAC 1747
 DB 1499 uAspAsnLeuIleProGluGlnIleAsn 1508

RESULT 8
 PCT-US94-09943-2
 ; Sequence 2, Application PC/TUS9409943
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; APPLICANT:
 ; APPLICANT:

APPLICANT: TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: TWONEY, MICHAEL J.
REGISTRATION NUMBER: P-38,349
REFERENCE/DOCKET NUMBER: LO461/7000WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
TELEX: 92-1742 EZEKIEL
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2466 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-09943-2

Alignment Scores:
Pred. No.: 4, 43e-09 Length: 2466
Score: 204.50 Matches: 143
Percent Similarity: 36.35% Conservative: 94
Best Local Similarity: 21.93% Mismatches: 214
Query Match: 5.13% Indels: 201
DB: 5 Gaps: 29

US-09-502-945-5 (1-2162) x PCT-US94-09943-2 (1-2466)

QY 1 CTGTGGCCGCTGGGCTGGGCTCTT-----27
Db 991 ProProGlnThrValAlaGluLeuValGlyLysProSerHisGlnMetSerArgSer 1010
QY 28 --TCCAGCTCTGGACCGCGCCGAGGAGGAGGCTGTCGCAACGACGAGCTGGA 84
Db 1011 AspaGluSerLeuAlaGlyValThrLysLeuAsnSerLysSerValAla-----1028
QY 85 CCTGGCCCATGGACCGCAAAAGTGGCCGAGAAATTCGGCATAGGTGGATTCTG 144
Db 1029 -----SerLeuAsnArgSerProGluArg---LysHisGluSerSerSer 1044
QY 145 ATTGAAATGATCAGAGAGGACTATCTATGATGCTGCGCAATGTACCAACGAC 204
Db 1045 SerileGluAspProGlyGlnAlaTyrValLeuAspValLeu-----HisLysArg 1061
QY 205 ATGACGTGGCGTCTGCTGGGAGACCTGAAGCTGGTCAATCAATCAACCGCGTCTG 264
Db 1062 TrpSer-----ileValSerSerProGluArg---1070
QY 265 CCTCTGTTGATGCATTCGCCCTGATCCCACTGAAGCAGCAGGTGGGAATATGATCAG 324
Db 1071 -----GluileThrLeuValAsnLeuLysAspAlaTyr-----1083

QY 325 CTGACCCCGCGCTCCAGGAAGCTGAAGGAGGTGCTCTGGACCGCTCTGCACCCCGAA 384
Db 1083 -----1083
QY 385 GGCCTCGGCTGAGTGTGGTGGCTGGAGTTTGGC-----TGTGGGCTCTTC 435
Db 1084 GlyLeuGlyPheGlnIleIleGlyGlyLeuMetGlyArgLeuAspLeuGlyIlePhe 1103
QY 436 ATCTCCACCTCATCAAAAGGCGTCAGGACAGACAGCTCGG-----CTCAGTAGGGAC 492
Db 1104 IleSerSerValAlaProGlyGlyProAlaAspPheHisGlyCysLeuLysProGlyAsp 1123
QY 493 GAGATCGCGGATCAATGATATTCATCTCTCTCTTACCCATGAGAGGTCTATCAAC 552
Db 1124 ArgLeuIleSerValAsnSerValSerLeuGluGlyValSerHisAlaIleGlu 1143
QY 553 CTCATTGCAACCAAGAAACTGTCTCCATCAAGTGAGACACATCGGCTGTATCCCGTG 612
Db 1144 IleLeu-----1145
QY 613 AAAAGCTCTCTGATGAGCCCTCCTTGGCAGATGTGGATCAGTTTGTGCGAATCT 672
Db 1146 GlnAsnAlaProGluAsp-----ValThrLeuValIleSerGlnPro 1159
QY 673 GGGGGGTGCGAGGAGCGCTGGCTCCCTCGAAATCGGGAACAGAGAGAGGTC 732
Db 1160 -----LysGluLys-----1162
QY 733 TTCATCAGCTGTAGTCTCCGAGGCTTGGCTGCAGCATTTCCAGCGGCCCATCCAG 792
Db 1163 -----LysValProSerThrProValHis 1172
QY 793 AAGCCTGGCATCTTTATCAGCCATGTGAACCTGGCTCCCTCTCTGCTGAGTGGGATTG 852
Db 1173 LeuThrAsnGluMetLysAsnTyrMetLysLysSer-----1185
QY 853 GAGATAGGGACCATGTCGAAAGTCAATGGCGTGGCTCTTCAACCTGATCACAAG 912
Db 1186 -----TyrMetGlnAspSerAlaIleAspSerSerSerLysAspHis 1200
QY 913 GAGGCTGTAAATGTCTGCAAAATAGCGCAGCTTGCATCTCCATCTCTAGCTCAGCT 972
Db 1201 TrpSerArgGlyThrLeu-----ArgHisIleSerGluAsnSerPheGlyProSer 1217
QY 973 GGC-----CGGAGCTGTTCATGACAGACCGGAGCGGCTGGCAGCGCGGAGCCT 1026
Db 1218 GlyGlyLeuArgGlySerLeuSerSerGlnAspSerArgThrGluSerAlaSerLe 1237
QY 1027 GAGCTGCAGC-----GGCAGGAGCTTCTCATGCGAAGCGGC-----TGGCGATG 1071
Db 1237 uSerGlnSerGlnValAsnGlyPhePheAlaSerHisLeuGlyAspGlnThrTrpGlnG 1257
QY 1072 GAGTCCAACAAGATCTCTCCAGGAGCAGGAGATGGAGCGGCAAGAAAGAAAT 1131
Db 1257 uSerGlnHisGlySerProSerProSerValIleSerLysAlaThrGluLysGluThrP 1277
QY 1132 GCCCAAGAGCAGCAGAGAAATAGAGATACCGGAA-----GGAGATGGAACAGATTGT- 1187
Db 1277 heThrAspSerAsnGlnSerLysThrLysLysProGlyIleSerAspValThrAspTyrS 1297
QY 1188 -----AGAGGGAAGAGAAATTAAGAGCAATGGGAAGAGACTGGGGCTCAAGGAA 1242
Db 1297 erAspArgGlyAspSerAspMetAspGluAla-----T 1308
QY 1243 GAGTACTCTTCCCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCCGACGCA 1302
Db 1308 hrTyrSerSerGlnAspHisGln-----ThrProLysGlnGluSerSerSers 1325
QY 1303 AAGTATTTCCGGAATATGAGGAAGGCTTTCACCCCTACTCTATGTTCCACCCGAGCA 1362
Db 1325 erValAsnThrSerAsn-----LysMetAsnPhelysThrPheSerSerProProLysPr 1344
QY 1363 GATCATGGGGAAGGATGTCCGCTCTCTACGCATCAAGAAGAGGAGGATCTTACACCTGC 1422

Db 377 ProLeuLysProProGluAlaValArgIle-GlnArgSerGluGlnLeuHisHisTrpAs 396
QY 1028 -----AGCTGC-----ACGGCAGGAGCTTCTCATGACAGCGCGTGGCGATG 1071
Db 396 pProCysValProSerCysHisSerProArgProSerHisCysArgAlaPro---ThrTr 415
QY 1072 GAGTCCA-----ACAAGATCCTCCAGGAGCAGCAGGAGATGAGCGGC-----1114
Db 415 pAlaProGlyGlyGlnAspGlnSerArgSerValSerSerThrProPheSerSerProTh 435
QY 1115 -----AAAGGAAAGAAAGAAATGTCAGAGAGCGCAGCAGAGAA 1152
Db 435 rMetAsnProAlaPheProCysAlaAsnAlaSerThrLeuProArgGlyProMetSerPr 455
QY 1153 AATGAGAGATACCGAAGGAGATGAGACAGATGTGAGGAGGAGAGA-----1201
Db 455 oArgThrThrAlaGlyArgArgGln-----ArgArgLysGluHisArgSerSe 472
QY 1202 ----AGTTTAAAGCAATGGGAAGAGACTGGGCTCAAGAGACAGCTACTCTCGCT 1257
Db 472 rLeuSerLeuAlaSerSerThrValGlyProGlyGlyGln-----485
QY 1258 AAAACCATCATGCTGAGGTACACCCAGTACCCCTCCCAAGCCAAAGTGTATTCGGA 1317
Db 486 -----IleValHisThrGlu-----490
QY 1318 ATATGAGGAAGCTTACCCCTACTATGTTACCCAGCAGAGATCATGGGAGGA 1377
Db 491 -----ThrThrGluValValLeuCysGlyAs 499
QY 1378 TGTCGGGCTCTACGCATCAAGAGGAGGATCTTAGACTGCGCTGGAAGCGGTGT 1437
Db 499 p-----ProLeuSerGlyPheGlyLeuGlnLeuGlnGlyGlyI 512
QY 1438 G-----GACTCCCCCATTTGGGAAGGTGCTGCTTCTGCTGTATGA 1479
Db 512 ePheAlaThrGluThrLeuSerPro-----ProLeuValArgPheIleGluPr 529
QY 1480 CGGGGAGCTCTGAGCGGATGTCGATGTCAAGGGGACAGATCATGGCAATCAA 1539
Db 529 oAspSerProAlaGluArgCysGlyLeuLeuGlnValGlyAspArgValLeuAlaIleAs 549
QY 1540 CGGCAAGATTGTGACAGATACACCTGGCTGAGGCT-----GACGCTGC 1584
Db 549 nGlyIleAlaThrGluAspGlyThrMetGluGlnAlaAsnGlnLeuLeuArgAspAlaI 569
QY 1585 CTTCAAGAGCCCTGGAATCAGCGGGGAGTGCATGATCGACTGTGTTGCTCCGCTGCC 1644
Db 569 aLeuAlaArgLysIleValLeuGluIleGluPhe---AspValAlaGluSerValIlePr 588
QY 1645 C-----CCAAAGGATGATGACCATGACATGAGCTGACCTTCTT 1677
Db 588 oSerSerGlyThrPheHisValLysLeuProLysArgArgGlyValGluLeuGlyIleTh 608
QY 1678 GCTGAAGTCC-----AAAAGGGAAACAAATTCACCGCTTAGGAACACAGTGAGCT 1728
Db 608 rIleSerSerAlaSerArgLysArgGly-----617
QY 1729 CCGGCCCCACCTCTGTAAACACAAAGCCTCGGACCGCTTGAGAGAGGCCACATGACACA 1788
Db 618 -GluProLeuIleIle-----SerAspIleLysLysGlySerValAlaIh 632
QY 1789 CACC-----AGATGGCATCTTGGGACCTGAATCTATCACCAGGAATCTCAACATC-- 1840
Db 632 sArgThrGlyThrLeuGluProGlyAspLysLeuLeuAlaIleAspAsnIleArgLeuAs 652
QY 1841 -----CCTTGGCC-----CTGAACACGAGG 1860
Db 652 pHisCysProMetGluThrAlaValGlnIleLeuProGlnCysGluAspLeuValLysLe 672
QY 1861 CCAGATGAAGAACAGCTCGGCCACTTTTGTGAAGGCAATGTGGAGCA-----1909
Db 672 uLysIleArgLysAspGluAspAsnSerAspGluGlnGluSerSerGlyAlaValSerTy 692

QY 1910 -----AAGGAGCAGCAGCCGTTTGGGA-----GAAGA 1938
Db 692 rThrValGluLeuLysArgTyrglyGlyProLeuGlyIleThrIleSerGlyThrGluGl 712
QY 1939 TCACAAGATCCAGACTCTCATTCCTTCTCTGCGCCAGTGAATTTGGTCTCTCCAGC 1998
Db 712 uProPheAspPro-IleIleIleSerGlyLeu-----ThrLysArgGlyLeuAlaGluA 730
QY 1999 TTTGGGAGACTCTCTTCCTGAACCTAATAAGACCCAGCTGCTCTCTCTCCATC 2058
Db 730 rgThrGlyAlaSerMetLeuGlyThrAlaTyrrPrProSer***SerValSerLeuLysG 750
QY 2059 CTTCTCTCTGCTGCTCTCTTAATGCTGCCAGGATGTCTCAAACTTACTCTG 2117
Db 750 lyArgProLeuSerGluAlaIleHisLeuLeuGlnValAlaGlyGluThrValThrLeu 769
RESULT 11
US-09-045-632-2
; Sequence 2, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: HUGANIR, RICHARD L.
; APPLICANT: DONG, HUALING
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1112 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-045-632-2

Alignment Scores:
Pred. No.: 4,64e-09 Length: 1112
Score: 202.00 Matches: 144
Percent Similarity: 35.93% Conservatives: 82
Best Local Similarity: 22.89% Mismatches: 203
Query Match: 5.06% Indels: 201
DB: 3 Gaps: 26

US-09-502-945-5 (1-2162) x US-09-045-632-2 (1-1112)

QY 322 CAGCTGACCCCGCGCTCCAGGAAGTGGCTGACCGTCTGACCGCTCTGACCC 381


```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1112 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-045-632-3

```

```

Alignment Scores:
Pred. No.: 4,648-09 Length: 1112
Score: 202.00 Matches: 144
Percent Similarity: 35.93% Conservative: 82
Best Local Similarity: 22.89% Mismatches: 203
Query Match: 5.06% Indels: 201
DB: 3 Gaps: 26

```

US-09-502-945-5 (1-2162) x US-09-045-632-3 (1-1112)

```

QY 322 CAGCTGACCCCGCGCTCCAGAGAGTGAAGAGGTCGCTGACCGCTGTCACCC 381
DB 42 GlnSerIleProGluGluPheLysGlySerThrValValGluLeuMetLysGluGly 61
QY 382 GAAGCCCTCGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 441
DB 62 ThrThrLeuGlyCysThrValSerGlyGlyIleAspLysAspGlyLysProArgValSer 81
QY 442 CACCTCATCAAGCGGT---CAGCGACAGACGCGCTCCAGTGGGACGAGATC 498
DB 82 AsnLeuArgGlnGlyIleAlaAlaArgSerGlnLeuAspValGlyAspTyrIle 101
QY 499 GTCCGGATCAATGGATATCCATCTCCTCTGATCCCATGAGGAGTGCATCAACCTCAT 558
DB 102 LysAlaValAsnGlyIleAsnLeuAlaLysPheArgHisAspGluIleSerLeuLeu 121
QY 559 CGA---ACCAAGAAACTGTGTCCCAAGTGAACATGACACATCGGCTGATCCCGTGA 615
DB 122 LysAsnValGlyGluArgValValLeuGluValGluTyr-----GluLeuProVal 139
QY 616 ACCTCTCTGATGACCCCTCACTGGCAGTATGGATCACTTGTGTGGAAATCTGGG 675
DB 140 SerIleGlnGlySerValMetPheArgThrValGluValThrLeuHisLysGluGly 159
QY 676 GCGGTGCGAGGACGCTG---GGCTCCCTCGGAATCGGAAACAGAGAGAGAG 729
DB 160 AsnThrPheGlyValIleArgGlyGlyAlaHisAspAspArgAsnLysSerArgPro 179
QY 730 GTCTTCATCAGCTGGTAGGCTCCGAGGCTTGGCTGCAGCATTTCCAGCGGCCCATC 789
DB 180 ValValIleThr-----CysValArgProGlyGlyProAsp 191
QY 790 CAGAAGCCTGGCATCTTTATCAGCATGTGAAACCTGGCTCCCTGCTGCTGAGTGGA 849
DB 192 AspArgGluGlyThr-----IleLysPro----- 199
QY 850 TTGGAGATAGGGGACAGATTCGAATGCAATGCGGTGCGACTTCTCAACCTGGATCAC 909

```

```

DB 200 -----GlyAspArgLeuLeuSerValAspGlyIleArgLeuLeuGlyThrThrHis 111
QY 910 AAGGAGCTGTAAATGTCTGAAA-----AATACCGCAGCTCACCATC----- 954
DB 217 AlaGluAlaMetSerIleLeuLysGlnCysGlyGlnGluAlaThrLeuLeuIleGluTyr 236
QY 955 -----TCCATTGTAGCTGCAGCTGGCCGG-----CAGCTGTTC 987
DB 237 AspValSerAlaMetAspSerValAlaThrAlaSerGlyProLeuLeuValGluValAla 256
QY 988 ATGACAGACCGGCGCTGGCAGCGCGG-----1019
DB 257 LysThr---ProGlyAlaSerLeuGlyValAlaLeuThrThrSerValCysCysAsnLysG 276
QY 1020 -----GCAGCGTGTAGCTGCAGCGCGCAGCGAGCTTCTC 1050
DB 276 InValIleValIleAspLysIleLysSerAlaSerIleAlaAspArgCysGlyAlaLeuH 296
QY 1051 ATGACAGACCGGCTGGCAGTGCATCAACAAGATCTCCAGGACGACGAGAGTGGAG 1110
DB 296 Is-----ValGlyAspHisIle-----LeuSerIleAspGlyT 307
QY 1111 CGGCAAGGAGAGAAATTCGCCAGAGGCGACGAGGAGAAATGACAGATACCGGAG 1170
DB 307 hrSerMetGluTyrCysThrLeuAlaGluAlaThrGlnPheLeu-----G 322
QY 1171 GAGATGGAACAGATTGTAGAGGAGGAGAGAGTGTAAAGAACATGGGAAGAGACTGG 1230
DB 322 LysAsnThrThrAsp---GlnValLysLeuGluIleLeuProHisGlnThrArgLeuA 341
QY 1231 GGTCTCAAAGGA-----1241
DB 341 laLeuLysGlyProAspHisValLysIleGlnArgSerAspArgGlnLeuProTrpAsp 361
QY 1242 -----ACAGCTACTCTTGCCTAAACCATCACTGCTGAGGTACACCCAGTA 1287
DB 361 roTrpAlaSerSerGlnCysSerValHisThrAsnHisHis-----HisAsnProHisH 379
QY 1288 CCCCTTCCCAAGCCAAAGTGTATTCGGGAAATATGAGGAGGCTTTGACCCCT----- 1340
DB 379 lsProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsnSerP 399
QY 1341 -----ACTCTATGTTCAACCCAGACAGATCATGGGAGAGGATGTCGGGCT 1386
DB 399 roProAlaMetValSerSerSerProThrSerMetSerAlaTyrSerLeuSerSerL 419
QY 1387 C-----1387
DB 419 euAsnMetGlyThrLeuProArgSerLeuTyrSerThrSerProArgGlyThrMetMetA 439
QY 1388 --TACGCGCATCAAGAAG-----GAGGATCTCTTAGACCTGCGC----- 1423
DB 439 rgArgArgLeuLysLysLysAspPheLysSerSerLeuSerLeuAlaSerSerThrValG 459
QY 1423 -----CTGGAAGCGCGGTGTG-----1423
DB 459 lyLeuAlaGlyGlnValValHisThrGluThrThrGluValValLeuThrAlaAspProv 479
QY 1424 -----CTGGAAGCGCGGTGTG-----GACTCCC 1445
DB 479 alThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThrGluThrLeuSerSerP 499
QY 1446 CCATTGGGAAGTGTGCTTCTGCTGTATGAGCGGGGAGCTGTGAGCGGATGGTG 1505
DB 499 ro-----ProLeuIleSerTyrIleGluAlaAspSerProAlaGluArgCysGlyV 516
QY 1506 GCATTGTGAAGGGGACGAGATCATGGCAATCAACGGCAGATGTGTGACAGATACACCC 1565
DB 516 alLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIleProThrGluAspSerThrP 536
QY 1566 TGGCTGAGGCTGAGCTGCCCTGCGAAGAGGCTGGAATCAGGCGGGGACTGGATC----- 1621

```

Db 536 heGluGluAlaAsnGlnLeuLeuArgAspSerSerIleThrSerLysValThrLeuGluI 556
QY 1622 -----GACCTTGTGGTCCGCTCCGCC-----C 1646
Db 556 leGluPheAspValAlaGluSerValIleProSerSerGlyThrPheHisValLysLeup 576
QY 1647 CAAAGGAGTATGACGATGACGTGACCTTCTGCTGAAGTCC----- 1687
Db 576 roLysLysHisSerValGluLeuGlyIleThrIleSerSerProSerSerArgLysProG 596
QY 1688 -----AAAAGGGAACCAATTCACCGCTTAGGAACA 1721
Db 596 lyAspProLeuValIleSerAspIleLysLysGlySerValAlaHisArgThrGlyThrL 616
QY 1722 GNGAGCTCGGGCCCACTCGG 1744
Db 616 euGluLeuGlyAspLysLeuLeu 623
RESULT 13
US-09-045-632-14
; Sequence 14, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Huguair, Richard L.
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-045-632-14
Alignment Scores:
Pred. No.: 4e-09 Length: 604
Score: 201.00 Matches: 142
Percent Similarity: 36.47% Conservative: 83
Best Local Similarity: 23.01% Mismatches: 190
Query Match: 5.04% Indels: 203
DB: 3 Gaps: 27
US-09-502-945-5 (1-2162) x US-09-045-632-14 (1-604)

QY 364 CTGACCGTCTGCACCCCGAAGGC-----CTCGGCCTGAGTGTGCTGTGCTCGCTGGAG 417
Db 3 ValGluLeuMetLysLysGluGlyThrThrLeuGlyCysThrValSerGlyGlyLeuAsp 22
QY 418 TTTCGGCTGTGGCTCTTCATCTCCACCTCATCAAGCCGGT---CAGGCAGACAGCGTC 474
Db 23 LysAspGlyLysProArgValSerAsnLeuArgGlnGlyIleAlaAlaArgSerAsp 42
QY 475 GGGCTCCAGGTAGGGGACGAGATCGTCGGGATCAATGATATTCATCTCTCTCTGCTGACC 534
Db 43 GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg 62
QY 535 CATGAGGAGTCAACCTCATTCGA---ACCAAGAAACTGTGTCCATCAAGTGAAGA 591
Db 63 HisAspGluIleLeuSerLeuLeuLysAsnValGlyGluArgValValLeuGluValGlu 82
QY 592 CACATCGGCCTGATCCCGGTGAAAAGCTCTCTCTGATGAGCCCTCACTTGGCAGTGTG 651
Db 83 Tyr-----GluLeuProValSerIleGlnGlySerSerValMetPheArgThrVal 100
QY 652 GATCAGTTTGTGCGGAATCTGGGGCTGGAGGCGAGCCTG-----GGCTCCCTCGGA 705
Db 101 GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyGlyAlaHis 120
QY 706 AATCGGAAACAAAGAGAGAGAGGTCTTCATCAGCCTGTGTAGGTCCCGAGGCTTGGC 765
Db 121 AspAspArgAsnLysSerArgProValValIleThr----- 132
QY 766 TGCAGCATTTCCAGCGGCCCATCCAGAACCTGGCATCTTTATCAGCCATGTGAACCT 825
Db 133 CysValArgProGlyGlyProAspAspArgGluGlyThr-----IleLysPro 148
QY 826 GGCTCCCTGTCTGCTGAGGTGGATTGGAGATAGGGGACACAGATTGTGCAAGTCAATGGC 885
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157
QY 886 GTCGACTTCTTAACCTGGATCACAGGAGCTGTAAATGTGCTGAA-----AAT 936
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177
QY 937 AGCCGACGCTGACCATC-----TCCATTGTAGTGCAGCT 972
Db 178 GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer 197
QY 973 GGCCGG-----GAGCTTTCATGACAGCGGGGAGCGGTGGCAGAGCGCG----- 1019
Db 198 GlyProLeuLeuValGluValAlaLysThr--ProGlyAlaSerLeuGlyValAlaLeu 217
QY 1020 -----GCAGCGT 1026
Db 217 hrThrSerValCysAsnLysGlnValIleValIleAspLysLysSerAlaSerI 237
QY 1027 GAGCTGCAGCGCAGGAGCTTCTCATGCAGAAGCGGCTGGCGATGGATGCCAACAGATC 1086
Db 237 leAlaAspArgCysGlyAlaLeuHis-----ValGlyAspHisIle----- 250
QY 1087 CTCAGGAGCAGCAGGAGATGGCGGCAAGAGAGAAAGAAATTCGCCAGAGACGACA 1146
Db 251 -----LeuSerIleAspGlyThrSerMetGluTyrCysThrLeuAlaGluAlaThrG 268
QY 1147 GAGGAAATGAGAGATACCGGAAGAGAGATGGAACAGATTTAGAGGAGAGAGAGTTT 1206
Db 268 InPheLeu-----GlyAsnThrThrAsp---GlnValLysLeuGluIleL 282
QY 1207 AAGAAGCAATGGGAAGAACTGGGGCTCAAGGA----- 1241
Db 282 euProHisHisGlnThrArgLeuAlaLeuLysGlyProAspHisValLysIleGlnArgS 302
QY 1242 -----ACAGCTACTCTCTCCCTAAACACC 1263
Db 302 erAspArgGlnLeuProTrpAspProTrpAlaSerSerGlnCysSerValHisThrAsnH 322
QY 1264 ATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTGTATTCGGAATATGA 1323

```

Db 322 1SHIS-----HISAsnProHisProAspHisCysArgValProAlaLeuGlyPhep 340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
*QY 1324 GAGAGCTTTGACCCCT-----ACTATATTTTACCCCGAGAGCA 1362
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 340 rolysalaleuThrProAsnSerProAlaMetValSerSerProThrSerM 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1363 GATCATGGGAAGGATCGGCTC-----1387
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 eSerAlaTyrSerLeuSerSerLeuAsnMetGlyThrLeuProArgSerLeuTyrSer 380
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1388 -----CTACGCATCAAGAAG-----GAGGGATCTT 1412
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 hSerProArgGlyThrMetMetArgArgLeuLysLysAspPheLysSerSerL 400
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1413 TAGACCTGCC-----1423
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 400 euserLeualaSerSerThrValGlyLeualaGlyGlnValValHisThrGluThrG 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1424 -----CTGGAAGCGGTGTG- 1438
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 luValValLeuThrAlaAspProValThrGlyPheGlyLeGlnLeuGlnGlySerValP 440
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1439 -----GACTCCCCCATTTGGGAAGGTGCTGCTTCTGCTGTATGAC 1481
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 440 heaLathrGluThrLeuSerSerPro-----ProLeuLeSerTyrIleGluAlaA 457
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1482 GGGAGCTCTGAGCGGATGTTGGATGTTGAAAGGGAGCAGATCATGGAATCAAG 1541
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 457 spSerProAlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetAlaIleAsnG 477
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1542 GCAAGATTGTGACAGACTACCCGTGGGTGAGGTGAGCTGCTGCTGAGAGGCTGGA 1601
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 477 lyileProthrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerSerI 497
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1602 ATCAGGGGGGACTGGATC-----GACCTTGTGTTCCGCTGCCCTC----- 1645
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 497 leThrSerLysValThrLeuGluLeuPheAspValAlaGluSerValIleProSerS 517
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1646 -----CCAAAGGATGATGACGATGACGTGACCTTCTTGTGCTGA 1682
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 517 exGlyThrPheHisValLysLeuProLysLysHisSerValGluLeuGlyIleThrIleS 537
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1683 ACTCC-----AAAAGGGGA 1697
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 537 exSerProSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlys 557
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1698 ACCAAATTCACCGTTAGGAACAGTACGCTCGGCGCCCACTCGTG 1744
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 557 erValAlaHisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 14
US-09-045-632-15
; Sequence 15, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Huanir, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045.632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 702 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-045-632-15

```

```

Alignment Scores:
Pred. No.: 4,36e-09 Length: 702
Score: 201.00 Matches: 142
Percent Similarity: 36.47% Conservative: 83
Best Local Similarity: 23.01% Mismatches: 190
Query Match: 5.04% Indels: 203
DB: 3 Gaps: 27

```

US-09-502-945-5 (1-2162) x US-09-045-632-15 (1-702)

```

QY 364 CTGGACCTCTGCACCCCGAAGGC-----CTCGSCCTGAGTGTGGTGGCTGGAG 417
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 ValGluLeuMetLysLysGluGlyThrLeuGlyCysThrValSerGlyGlyLeasp 22
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 418 TTTGGCTGGGCTCTTCATCTCCACCTCATCAAGCGGT---CAGGCAGACGCTC 474
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23 LysaspGlyLysProArgValSerAsnLeuArgGlnGlyGlyIleAlaAlaArgSerAsp 42
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 475 GGGTCCAGGTAGGGGACGAGATGTCGGGATCATGATATTCATCTCTCTCTGTACC 534
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43 GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPhearg 62
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 535 CATGAGGAGGTCACTCAACTCATTCGA---ACCAAGAAACTGTCTCATCAAGTGA 591
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 HisaspGluIleIleSerLeuLysAsnValGlyGluArgValValLeuGluValGlu 82
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 592 CACATCGGCTGATCCCGTGAAGAGTCTCTCTGATGAGCCCTCACTTGGCAGTATGTG 651
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83 Tyr-----GluLeuProProValSerIleGlnGlySerSerValMetPheArgThrVal 100
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 652 GATCAGTTTGTGCGGATCTGGGGCGTGGAGCAGCCTG-----GGTCCCTGGA 705
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101 GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyAlaHis 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 706 AATCGGGAACAAGAGGAAGGTCTTCATCAGCCTGTTAGCTCCCGAGCCTTGGC 765
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 AspArgAsnLysSerArgProValIleThr-----132
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 766 TCGAGCATTTCCAGCGCCCATCCAGAGCCTGGCATCTTTATCAGCATGTGAACCT 825
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133 CysValArgProGlyGlyProAspArgGluGlyThr-----IleLysPro 148
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 826 GGCTCCCTGTCTGCTGAGGTGGGATTTGGAGATGGGACACAGATTGCAAGTCAATGGC 885
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 886 GTCGACTTCTTAACCTGGATCACAAGGAGCTCTTAATGTGCTGAAA-----AAT 936
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```
QY 475 GGCCTCCAGGTAGGAGGAGATCGTCGGGATCAATGGATATTTCATCTCCCTCTCTGACC 534
    ||| ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 43 GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg 62

QY 535 CATGAGAGGTCAATCAACCTCATTCGA---ACCAAGAAACTGTGTCCATCAAGTGA 591
    ||| ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 63 HisAspGluIleSerLeuLysAsnValGlyGluArgValValLeuGluValGlu 82

QY 592 CACATCGGCTGATCCCGCTGAAGAACTCTCTGTAGTACGCCCTCACTTGGCAGTATGG 651
    :: ::||| ||| ||| ::||| ||| ::||| |||
Db 83 Tyr-----GlnLeuProProValSerIleGlnGlySerValMetPheArgThrVal 100

QY 652 GATCAGTTGTGTGCGGAATCTGGGCGCTCGAGGAGCAGCTG-----GCCTCCCTGGA 705
    :: ::||| ||| ||| ::||| ||| ::||| |||
Db 101 GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyGlyAlaHis 120

QY 706 AATCGGGAACAAAGAGAGAGGTCTTCATCAGCAGCTGTGGATCTTTATCAGCCATGTGAACCT 825
    :: ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 121 AspAspArgAsnLysSerArgProValValIleThr----- 132

QY 766 TGCAGCATTTCCAGCGGCCCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAACCT 825
    ||| ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 133 CysValArgProGlyGlyProAspAspArgGluGlyThr-----IleLysPro 148

QY 826 GGCTCCCTCTGCTGAGGTGGGATTGGAGATAGGAGGACCAGATTGTCCAGATCAATGGC 885
    ||| ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157

QY 886 GTGCACTTCTTAACCTGGATCACAGGAGGTGTAATGTGCTGAAA-----AAT 936
    :: ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177

QY 937 ACCCGCAGCTGACCATC-----TCCATTGACTGCAGCT 972
    ::||| ||| ||||| ||| ::||| ||| ::||| |||
Db 178 GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer 197

QY 973 GGCCGG-----GAGCTGTTTCATGACAGCCGCGGCTGCGCAGGCGCG--- 1019
    ||| ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 198 GlyProLeuLeuValGluValAlaLysThr--ProGlyAlaSerLeuGlyValAlaLeuT 217

QY 1020 -----GCAGCGT 1026
    ||| ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 217 hrThrSerValCysAsnLysGlnValIleValIleAspLysIleLysSerAlaSerI 237

QY 1027 GAGCTCCAGCGCAGGAGTCTTCATGACAGAGCGCTGCGATGGATGCCAACAGATC 1086
    ||| ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 237 leAlaAspArgCysGlyAlaLeuHis-----ValGlyAspHisIle----- 250

QY 1087 CTCAGGAGCAGGAGATGGAGCGGCAAGAGGAAAGAAATTCGCCAGAGGAGCGCA 1146
    ::||| ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 251 -----LeuSerIleAspGlyThrSerMetGluTyrCysThrLeuAlaGluAlaThrG 268

QY 1147 GAGGAAATGAGAGATACCGGAGGAGATGGAACAGATTGTAGAGGAGGAGGAGGTTT 1206
    :: ||||| ||| ::||| ||| ::||| ||| ::||| |||
Db 268 InPheLeu-----GlyAsnThrThrAsp---GlnValLysLeuGluIleL 282

QY 1207 AAGAAGCATGGGAAGACACTGGGCTCAAGGA----- 1241
    ||||| ||||| ||||| |||||
Db 282 euProHisHisGlnThrArgLeuAlaLeuLysGlyProAspHisValLysIleGlnArgS 302

QY 1242 -----ACGCTACTCTTGCTTAAACCC 1263
    ::||| ||||| ||||| |||||
Db 302 erAspArgGlnLeuProTrpAspProTrpAlaSerSerGlnCysSerValHisThrAsnH 322

QY 1264 ATCACTGTGAGGTACACCCATACCCCTTCGCAAGCCCAAGTATTTCCGGAAATATGA 1323
    ||||| ||||| ||||| |||||
Db 322 ishHis-----HisAsnProHisHisProAspHisCysArgValProAlaLeuGlyPheP 340

QY 1324 GGAAGCCTTTGACCCCT-----ACTATGTTTACCCCGAGCA 1362
    ||||| ||||| ||||| |||||
Db 340 roLysAlaLeuThrProAsnSerProProAlaMetValSerSerSerProThrSerM 360
```

```
QY 1363 GATCATGGGAAGGATGTCCGGCTC----- 1387
    ||| ::||| ||| ::||| ||| ::||| |||
Db 360 etSerAlaTyrSerLeuSerSerLeuAsnMetGlyThrLeuProArgSerLeuTyrSerI 380

QY 1388 -----CTACGCATCAAGAAG-----GAGGATCCT 1412
    ||||| ||||| ||||| |||||
Db 380 hrSerProArgGlyThrMetMetArgArgLeuLysLysLysAspPheLysSerSerL 400

QY 1413 TAGACCTGGCC----- 1423
    || |||||
Db 400 euSerLeuAlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrThrG 420

QY 1424 -----CTGAAGCGCGTGTG- 1438
    ||||| ||||| ||||| |||||
Db 420 luValValLeuThrAlaAspProValThrGlyPheGlyIleGlnLeuGlnGlySerValP 440

QY 1439 -----GACTCCCCCATGGGAAGGTGGTCTTCTGCTGTGTATGAGC 1481
    ||||| ||||| ||||| |||||
Db 440 heAlaThrGluThrLeuSerSerPro-----ProLeuIleSerTyrIleGluAlaA 457

QY 1482 GGGGAGCTGCTGAGCGGCATGGTGCATTTGTAAAGGGGAGAGATCATGCAATCAACG 1541
    ||||| ||||| ||||| |||||
Db 457 spSerProAlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetalIleAsnG 477

QY 1542 GCAAGATTGTGACACACTACACCTGGCTGAGGCTGAGGCTGCCCTGCAGAGGCTCGA 1601
    || ||||| ||| ||||| |||
Db 477 lyIleProThrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerSerI 497

QY 1602 ATCAGGGCGGGACTGGATC-----GACCTTGTGGTTCCTGCTGCCCC----- 1645
    ::||| ||||| ||||| |||||
Db 497 leThrSerLysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerS 517

QY 1646 -----CCAAAGGAGTATGACGATGACGTGACCTTCTTGTCTGA 1682
    ||||| ||||| ||||| |||||
Db 517 erGlyThrPheHisValLysLeuProLysLysHisSerValGluLeuGlyIleThrIleS 537

QY 1683 AGTCC-----AAAAGGGGAA 1697
    ||| ||||| ||||| |||||
Db 537 erSerProSerSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlyS 557

QY 1698 ACCAAATTCACGCGTTAGGAACACAGTACGCTCCGCGCCCGCCACCTCGT 1744
    :: ||| ||||| ||||| |||||
Db 557 erValAlaHisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572
```

Search completed: March 21, 2003, 13:09:23

Job time : 57.7489 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 13:00:02 ; Search time 23.1498 Seconds
(without alignments)
9985.861 Million cell updates/sec

Title: US-09-502-945-5
Perfect score: 3989
Sequence: 1 cctggccgcggcggtgcg.....tattttccagcttaaaaaa 2162

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 442306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPFC-spool/US09502945/runat_14032003_101101_19301/app_query.fasta_1.10979
-DB=Published_Applications_AA -OFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09502945 -CGN_1_1_60 -runat_14032003_101101_19301
-NCPU=6 -ICPU=3 -NO_XLPUX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:*
2: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:*
7: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pcp:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1175.5	29.5	268	10	US-09-739-907-114
2	435	10.9	120	9	US-09-925-299-778
3	435	10.9	120	10	US-09-925-299-778
4	417	10.5	98	10	US-09-739-907-60

5	417	10.5	113	10	US-09-739-907-119	Sequence 119, App
6	406	10.2	95	10	US-09-739-907-111	Sequence 111, App
7	318.5	8.0	260	9	US-10-078-090-151	Sequence 151, App
8	263	6.6	2037	9	US-09-951-402-3	Sequence 3, Appli
9	263	6.6	2037	10	US-09-951-401-3	Sequence 3, Appli
10	263	6.6	2037	10	US-09-922-101-3	Sequence 3, Appli
11	250.5	6.3	1881	9	US-09-998-425-3	Sequence 3, Appli
12	250.5	6.3	1881	9	US-09-997-977-3	Sequence 3, Appli
13	237	5.9	46	10	US-09-864-761-39112	Sequence 39112, A
14	234.5	5.9	767	10	US-09-919-497-59	Sequence 59, Appl
15	232	5.8	45	10	US-09-864-761-39113	Sequence 39113, A
16	227	5.7	47	10	US-09-864-761-49113	Sequence 49113, A
17	225	5.6	1736	10	US-09-919-497-98	Sequence 98, Appl
18	204.5	5.1	2485	10	US-09-802-669-46	Sequence 46, Appl
19	194	4.9	744	10	US-09-862-179A-1	Sequence 1, Appli
20	190	4.8	1337	10	US-09-757-781-62	Sequence 62, Appl
21	184	4.6	358	9	US-10-274-968-4	Sequence 4, Appli
22	184	4.6	358	10	US-09-823-356-6	Sequence 6, Appli
23	184	4.6	358	10	US-09-740-027-4	Sequence 4, Appli
24	183.5	4.6	1356	10	US-09-757-781-2	Sequence 2, Appli
25	181	4.5	434	12	US-10-001-843-179	Sequence 179, App
26	178	4.5	632	9	US-09-992-598-219	Sequence 219, App
27	178	4.5	632	9	US-09-989-293A-219	Sequence 219, App
28	178	4.5	632	9	US-10-063-547-40	Sequence 40, Appl
29	178	4.5	632	9	US-09-989-735-219	Sequence 219, App
30	178	4.5	632	9	US-09-990-444-219	Sequence 219, App
31	178	4.5	632	9	US-09-989-730-219	Sequence 219, App
32	178	4.5	632	9	US-09-990-436-219	Sequence 219, App
33	178	4.5	632	9	US-09-991-181-219	Sequence 219, App
34	178	4.5	632	9	US-09-993-687-219	Sequence 219, App
35	178	4.5	632	9	US-09-989-734-219	Sequence 219, App
36	178	4.5	632	9	US-09-997-653-219	Sequence 219, App
37	178	4.5	632	9	US-10-174-590-166	Sequence 166, App
38	178	4.5	632	9	US-10-176-758-166	Sequence 166, App
39	178	4.5	632	9	US-10-063-616-40	Sequence 40, Appl
40	178	4.5	632	9	US-10-175-737-166	Sequence 166, App
41	178	4.5	632	9	US-09-993-667-219	Sequence 219, App
42	178	4.5	632	9	US-10-063-502-40	Sequence 40, Appl
43	178	4.5	632	9	US-10-173-706-166	Sequence 166, App
44	178	4.5	632	9	US-10-175-738-166	Sequence 166, App
45	178	4.5	632	9	US-10-175-752-166	Sequence 166, App

ALIGNMENTS

RESULT 1
US-09-739-907-114
; Sequence 114, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; CURRENT FILING DATE: 2000-12-20
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,457
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 114
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-907-114

Alignment Scores:
Pred. No.: 4,75e-81 Length: 268
Score: 1175.50 Matches: 236
Percent Similarity: 90.11% Conservative: 1
Best Local Similarity: 89.73% Mismatches: 0
Query Match: 29.47% Indels: 26
DB: 10 Gaps: 1

US-09-502-945-5 (1-2162) x US-09-739-907-114 (1-268)

QY 964 GCTGAGCTGCGCGGAGCTGTTCTATGACAGCCGGGAGCGCTGGCAGAGCGCGGCAG 1023
Db 6 ALaAlaAlaGlyArgGluLeuPheMetThrAspArgGluArgLeuAlaGluAlaArgIn 25

QY 1024 CQTGAGCTGCAGCGCAGAGCTTCTCATGCAAGAGCGCTGGCGCATGAGTCCAAACAG 1083
Db 26 ArgGluLeuInArgGluInGluLeuMetGlnLysArgLeuAlaMetGlnSerAsnLys 45

QY 1084 ATCTCCAGGACGACGAGAGATGGAGCGGCAAGAGAGAAATGCCCCAGAGCCA 1143
Db 46 IleLeuGlnGluGlnGlnMetGluArgGlnArgArgLysGluIleAlaGlnLysAla 65

QY 1144 GCAGAGAAATGAGAGATACCGGAGGAGATGGACAGATTGTAGAGGAGGAAGAG 1203
Db 66 AlaGluGluAsnGluArgTyrArgLysGluMetGluGlnIleValGluGluGluLys 85

QY 1204 TTTAAGAAGCAATGGGAAGACACTGGGGCTCAAAGGAACAGCTACTCTTGCTTAAACC 1263
Db 86 PheLysLysGlnTrpGluLysPrpGlySerLysGluGlnLeuLeuLeuProLysThr 105

QY 1264 ATCACTGCTGAGTACACCCAGTACCCCTTCGCAAGCCA----- 1302
Db 106 IleThrAlaGluValHisProValProLeuArgLysProLysTyrAspGlnGlyValGlu 125

QY 1303 -----AAGTGA 1308
Db 126 ProGluLeuGluProAlaAspLeuAspGlyGlyThrGluGluGlnGlyGln-As 145

QY 1309 TTTCCGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTTTCAACCCAGAGCATCAT 1368
Db 145 PheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGluGlnIle 165

QY 1369 GGGGAAGATGTCGCGCTCCTACGCATCAAGAGGAGGATCCTTAGACCTGGCCCTGA 1428
Db 165 tGlyLysAspValArgLeuLeuArgIleLysLysGluLysSerLeuAspLeuAlaLeuG 185

QY 1429 AGCGGTGTGGACTCCCATTTGGGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1488
Db 185 uGlyGlyValAspSerProIleGlyLysValValSerAlaValTyrGluArgGlyAl 205

QY 1489 TGCTGACGCGCATGTGTGGCATTGTGAAGGGGACGAGATCATGGCAATCAACGGCAAGAT 1548
Db 205 aAlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLys 225

QY 1549 TGTGACAGACTACACCTGGCTGAGGCTGACGCTGCCCTGCAAGAGCCTGGAATCAGG 1608
Db 225 eValThrAspTyrThrLeuAlaGluAlaAspAlaAlaLeuGlnLysAlaTrpAsnGln 245

QY 1609 CGSGGAGTGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
Db 245 yGlyAspTrpIleAspLeuValValAlaValCysProProLysGluTyrAspAspGlu 265

QY 1669 GACCTTC 1675
Db 265 uThrPhe 267

RESULT 2
US-09-925-299-778
; Sequence 778, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 778
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (81)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-778

Alignment Scores:
Pred. No.: 4,89e-25 Length: 120
Score: 435.00 Matches: 97
Percent Similarity: 80.65% Conservative: 3
Best Local Similarity: 78.23% Mismatches: 14
Query Match: 10.90% Indels: 10
DB: 9 Gaps: 4

US-09-502-945-5 (1-2162) x US-09-925-299-778 (1-120)

QY 1196 AAGAGAGTTTAAAG---AGCAATGGGAAGAGACTGGGCTCAAGAGACAGACTACTCT 1252
Db 6 GluArgHisLeuArgValThrGlnGlyLeuGlyThrGlyAlaPheLeuGlyGly----- 23

QY 1253 TGCCTAAACCACTCACCTGCTGAGCTACACCCAGTACCCCTTCGCAAGCCAAAGTGATTTC 1312
Db 24 ---LeuArgProValLeu-----GlnProArgGlnGlyGln---AspPhe 36

QY 1313 CGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTTTACCCAGAGCATCATGGGG 1372
Db 37 ArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGluGlnIleMetGly 56

QY 1373 AAGATGTCGCGCTCCTACGCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAAGGC 1432
Db 57 LysAspValArgLeuLeuArgIleLysGluGlySerLeuAspLeuAlaLeuGluGly 76

QY 1433 GGTGTGGACTCCCATTTGGGAAGGTGCTGTTCTGCTGTGTATGACGGGGAGCTGCT 1492
Db 77 GlyValAspSer**IleGlyLysValValSerAlaValTyrGluArgGlyAlaAla 96

QY 1493 GAGCGCATGTGCGCATTTGAAAGGGGACGAGATCATGGCAATCAACGGCAAGTGTG 1552
Db 97 GluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLysIleVal 116

QY 1553 ACAGACTACACC 1564
Db 117 ThrAspTyrThr 120

RESULT 3
US-09-925-299-778
; Sequence 778, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0

Alignment Scores:

US-09-502-945-5 (1-2162) x US-09-739-907-60 (1-98)

US-09-502-945-5 (1-2162) x US-09-739-907-60 (1-98)

QY 736 ATCAGCCTGGTAGGCTCCCGAGGCCCTTGGCTGCAGCATTTCCAGCGGCCCATCCAGAAG 795

Db 30 IleSerLeuValGlySerArgGlyLeuGlyCysSerIleSerSerGlyProIleGlnLys 49

QY 796 CCTGGCATCTTATCAGCCATGTGAACCTGGCTCCCTGCTGCTGAGGTGGGATTGGAG 855

Db 50 ProGlyIlePheIleSerHisValLysProGlySerLeuSerAlaGluValGlyLeuGlu 69

856 ATAGGGGACCAGATTGTCTCGAAGTCAATGGCGTTCGACTTCTCTTAACCTGGATCACCAAGGAG 915

DD	70	iregiyaspqinilevagiuguvaiashnqi yvarasppmeserfasmdeuaspnisl ysgu 89
QV	916	SCTGTAAATGCTCTGAAAAATAGSCGCAGCCTGACCATCTCCATTGTAGCTGCAGCTGGC 975

Db 90 -Leu-----GlnLeuAl 93

QY 976 C G G G A G C T G T T C A 988

Db 93 aGlySerCysSer 97

RESULT 5
US-09-739-907-119

; Sequence 119, Application US/09739907
; Patent No. US20010012889A1

; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE: INFORMATION ON THE

; TITLE OF INVENTION: 36 Human secreted proteins
 ; FILE REFERENCE: PZ022P1
 ; CURRENT APPLICATION NUMBER: US 00/730 907

/ CURRENT AFFILIATION NUMBER: 03/03/139,901
 / CURRENT FILING DATE: 2000-12-20
 : PRIOR APPLICATION NUMBER: 09/348,457

; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567

; PRIOR FILING DATE: 1998-01-07
 ; PRIOR APPLICATION NUMBER: 60/070,692

; PRIOR FILING DATE: 1998-01-07
 ; PRIOR APPLICATION NUMBER: 60/070,704
 ; PRIOR FILING DATE: 1998-01-07

; PRIOR FILING DATE: 1998-01-07
 ; PRIOR APPLICATION NUMBER: 60/070,658
 ; PRIOR FILING DATE: 1998-01-07

```

; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 119
; LENGTH: 113

```

```

; TYPE: pr1
; ORGANISM: Homo sapiens
ms-09-730-907-110

```

Alignment Scores:

Pred. No.:	1.11e-23	Length:	113
Score:	417.00	Matches:	88

Percent Similarity: 83.81% Conservative: 0

Best Local Similarity: 83.81% Mismatches: 0
Query Match: 10.45% Indels: 17
DB: 10 Gaps: 1

US-09-502-945-5 (1-2162) x US-09-739-907-119 (1-113)

Qy 676 GCGCTGCGAGGAGCTGGCTCCCTGGAATCGGAAACAAAGGAGAAAGTCTTC 735
Db 26 GlyValArgGlySerLeuGlySerProGlyAsnArgGluAsnLysGluLysValPhe 45
Qy 736 ATGACCTGGTAGGCTCCCGAGGCTTGGCTGCAGCATTTCCAGGGCCCATCCAGAG 795
Db 46 IleSerLeuValGlySerArgGlyLeuGlyCysSerIleSerSerGlyProIleGlnLys 65
Qy 796 CTTGCACTTTATCAGCCATGTGAACCTGGCTCCCTGTCTGCTGAGTGGGATGGAG 855
Db 66 ProGlyIlePheIleSerHisValLysProGlySerLeuSerAlaGluValGlyLeuGlu 85
Qy 856 ATAGGGACCAAGATTGTCGAAGTCAATCGCTGCATCTCTTAACCTGGATCAACAAGGAG 915
Db 86 IleGlyAspGlnIleValGluValAsnGlyValAspPheSerAsnLeuAspHisLysGlu 105
Qy 916 GCTGTAATGTCTGCAAAATAGCGCAGCTGACCATCTCCATTGTAGTCCAGCTGCC 975
Db 106 -Leu-----GlnLeuAl 109
Qy 976 CGGAGCTGTTC 988
Db 109 aclySerCysSer 113

RESULT 6

US-09-739-907-111
; Sequence 111, Application US/09739907
; Patent No. US2001001289A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 111
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-907-111

Alignment Scores:
Pred. No.: 7, 02e-23 Length: 95
Score: 406.00 Matches: 90
Percent Similarity: 60.13% Conservative: 2
Best Local Similarity: 58.82% Mismatches: 1
Query Match: 10.18% Indels: 60
DB: 10 Gaps: 1

US-09-502-945-5 (1-2162) x US-09-739-907-111 (1-95)

Qy 1678 GCTAGAGTCAAAAGGGGAAACCAATTCACGGTTAGGAACAGTAGCTCGGCCCA 1737
Db 3 SerGluValGlnLysGlyLysProAsnSer-ProLeuGlyAsnSerGluLeuArgProHi 22
Qy 1738 CCTCGTGAACACAAAGCCCTCGGACCGCTTGAGAGAGGCCACATGACACACACAGATG 1797

Db 22 sLeuValAsnThrLysProArgThrSerLeuGluArgGlyHis----- 36
Qy 1798 GCATCCTTGGGACCTGAATCTATCACCAGGAATCTCAACTCCCTTTGGCCCTGAACCA 1857
Db 36 ----- 36
Qy 1858 GGGCCAGATAAGGAACAGCTCGGGCCACTTTTTTTGAAGGCAATGTGGAGAAAGGAGC 1917
Db 36 ----- 36
Qy 1918 AGCCAGCCGTTTGGGAGAGAGATCTCAAGGATCCAGACTCTCATCTCTTCTCTGGCCCA 1977
Db 37 -----Thr-IleProPheLeuTrippros 44
Qy 1978 GTGAATTTGGTCTCTCCAGCTTTGGGGGACTCTCTCTTGAACCCCTTAATAAGACCCAC 2037
Db 44 erGluPheGlyLeuSerGlnLeuTrpGlyThrProSerLeuAsnProAsnLysThrProL 64
Qy 2038 TGGAGTCTCTCTCTCCATCCCTCTCTCTGCGCTCTGCTCTAAATGCTGCCAGGATG 2097
Db 64 euGluSerLeuHisProSerProLeuProSerAlaLeuIleAlaAlaArgilev 84
Qy 2098 TCACCTCCAAACCTTACTCTGAGCTCATTAATAAA 2132
Db 84 alThrProAsnLeuThrLeuSerSerLeuIleLys 95

RESULT 7

US-10-078-090-151
; Sequence 151, Application US/10078090
; Publication No. US20030044815A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karra, Kalpana
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and
; FILE REFERENCE: DEX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 151
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-078-090-151

Alignment Scores:
Pred. No.: 4, 64e-16 Length: 260
Score: 318.50 Matches: 85
Percent Similarity: 52.32% Conservative: 39
Best Local Similarity: 35.86% Mismatches: 84
Query Match: 7.98% Indels: 29
DB: 9 Gaps: 6

US-09-502-945-5 (1-2162) x US-10-078-090-151 (1-260)

Qy 358 GTGCTCTGGACCGTCTGCACCCCGAGGCTCGCTGAGTGTGCTGTGGCTGGAG 417
Db 1 LeuArgArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleArgGlyGlySerGlu 19
Qy 418 TTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGGCGCTCAGGACAGACGCTCGG 477
Db 20 HisGlyValGlyIleTyValserLeuValGluProGlySerLeuAlaGluLysGluGly 39
Qy 478 CTCAGGTAGGGGACGAGATCGTCGGGATCAATGGATATTCCTCTCTCTCTGATCCAT 537

```
Db 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 538 GAGGAGGTGCATCACTTCATGCAACCAAGAACTGTGTCATCAAAAGTGAGACACATC 597
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 AlaGluAlaValLysAlaLeuLysGlySerLysLeuValLeuSerValTySerAla 79
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 598 GGCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTCCTCTGGCAGATGTGAT--- 554
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 80 GlyArgIleProGlyTyThrValThrAsnHisIleTyThrTrp-----ValAspPro 97
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 655 -----CAGTTTGTGCGAATCTGGGGCTCGGAGCGAGCGCTGGC----- 696
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 98 GlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg 117
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 697 ---TCCCTTGGAAATCGGAAAC-----AAGGAGAGAAG 729
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 118 GlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLys 137
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 730 GTCTTCATCAGCTGTAGCTCCGAGCGCTGGCTGCGCATTTCCAGCGCGCCATC 789
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 790 CAGAGGCTGGCATCTTTATCAGCATGTGAAACCTGGCTCCCTGCTGCTGAGGTGGA 849
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 158 TyrGlyLeuGlyIleTyThrGlyValAspProGlySerGluAlaGluGlySerGly 177
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 850 TTGGAGATAGGAGGACAGATTCGAAATCGCGTGCAGTCTCTTAACCTGGATCAC 909
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 910 AAGGAGGCTGTAATGTGCTGAAATAGCCAGCGCTGACCATCTCCATTGTAGCTGA 969
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThrVal----- 214
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 970 GCTGGCGCGGAGCTGTTCATGACAGCGGAGCGGCTGGCAGAGCGCGG 1020
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 215 -----LysaspValGlyArgLeuProHisAlaArg 224
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
RESULT 8
US-09-951-402-3
; Sequence 3, Application US/09951402
; Patent No. US20020168752A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/951,402
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 09/306,998
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 60/084,740
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-951-402-3
```

```
Alignment Scores:
Pred. No.: 1.79e-11 Length: 2037
Score: 263.00 Matches: 146
Percent Similarity: 36.47% Conservative: 75
Best Local Similarity: 24.09% Mismatches: 178
Query Match: 6.59% Indels: 207
DB: 9 Gaps: 26
```

US-09-502-945-5 (1-2162) x US-09-951-402-3 (1-2037)

QY 11. TCGCGGTGCGGCTCTTCCAGCTCTGCGACCGCGGACCCGAA-----GGAAAG 61

```
Db 1552 SerGlnSerLeuMetValProGlnSerGlySerProGluProGluSerIleArgAsnThr 1571
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 GGTCTGCAACGACGACGCTGCGCCAGCCATGGACGAAAGTGCCTGCGAGAAAT 121
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1572 SerArgSerSerThrProAlaIlePheAla-Ser---AspProAlaThrCysProIleIl 1590
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 122 TCCGCGCATAAAGTGATTTCTGATTGAAATGATGC----- 158
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1590 eProgly-----CysGluThrThrIleGluIleSerLys 1601
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 159 ---AGAGAAGACCTATCTCTATGATGTGCTG----- 186
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1601 sGlyArgThrGlyLeuGlyLeu-SerIleValGlyGlySerAspThrLeuLeuGlyAlaI 1621
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 187 -----CGAATGTACCACCAG-----ACCATGAGCGTGGCGCTCGCTGG 226
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1621 leIleIleHisGluValTyThrGluGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaG 1641
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 227 GAGACCTGAAGCTGTGCATCAAT----- 249
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1641 LyspGlnIleLeuGluValAsnGlyIleAspLeuArgLysAlaThrHisAspGluAlaI 1661
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 250 -----GAACCCAGCGCTGCTGCTCTGCTCTGTTGATGCCATTCGCGCTGA 292
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1661 leAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyArgAspGluA 1680
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 293 TCCCACTGAAGCAGCAGCTGGAATATGATCAGCTGACCCCGCGCTCCAGGAAGTGA 352
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1680 laProTyLysGluGluGluValCysAspThrLeuThr----- 1692
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 353 AGGAGGTGCTGTGACCGCTGTCACCCCGCGCTGCGCTGAGTGTGCTGCTGCC 412
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1693 ---IleGluLeuGlnLysLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA 1711
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 413 TGGAGTTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGAGCGGTGAGCAGACAG 472
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1711 rGAsn---AspThrGlyValPheValSerAspIleValLysGlyIleAlaAspAlaA 1730
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 473 TCGGG---CTCCAGTGGGAGCAGATCGTCCGATCAATGATATTCATCTCCTCT 529
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1730 spGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGlyGluAspValArgAsnA 1750
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 530 GTACCATGAGGAGGTGCATCAACCTCATTCGA---ACCAAGAAACTGTGTCCATCAAG 586
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1750 laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGlyThrValThrLeuGluV 1770
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 587 TGACACATCGGCTGATCCCGTGAAGAGCTCT-----CCTGATGAG----- 630
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1770 alGlyArgIleLysAlaGlyProPheHisSerGluArgArgProSerGlnSerSerGlnV 1790
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 631 -----CCCTCACCTGGCAGTATGTGGATCAGT 658
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1790 alSerGluGlySerLeuSerSerPheThrPheProLeuSerGlySerSerThrSerGluS 1810
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 659 TTGTGTCGGAATCTGGGGCGTGGAGGACGCTGGGCTCC-----CCTGGAATCGGG 712
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1810 erLeu---GluSerSerLysLysAsnAlaLeuAlaSerGluIleGlnGlyLeuArgT 1829
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 713 AAACAAGGAGAGAAGGCTCTTCATCAGCTGGTAGGCTCCGAGGCTGGCTGCACGA 772
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1829 hrValGluMetLysLys-----GlyProThrAspSerLeuGlyIleSerI 1844
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 773 TTTCAGCGG-----CCCATCCAGAGCGCTGGCATCTTTATCAGCCATGTGAAC 823
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1844 leAlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetHetiSP 1864
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 824 CTGGCTCCCTGCTGCTGAGGTG---GGATTGGAGATAGGGGACAGAGTGTTCGAAGTCA 880
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1864 roThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleC 1884
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 881 ATGGGTGCTACTTCTTAACCTGGATCAAGAGGAGGTGTAATGTGCTGAAAATAGCC 940
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```


Qy	773	TTTCCAGCGC-----CCCATCCAGAAGCTGGCATCTTTATCAGCCATGTGAAC	823
Db	1844	leAlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetMethIsp	1864
Qy	824	CTGGCTCCCTGTCGTCTGAGTG- --GGATTGGAGATAGGGACCACGATTGTCCAAGTCA	880
Db	1864	roThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgileValThrIleC	1884
Qy	881	ATGGCTCGACTTCTCTAACCTGGATCACAGAGAGGCTGTAATGCTGTAATAATAGCC	940
Db	1884	ysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValAsnLeuLeuLysAsnAla-	1903
Qy	941	GCAGCTTGACCATCTCCATTGTAGCTGCGAGCTGGCGGAGCTGTTTCATGACACACCGG	1000
Db	1904	--SerGlySerIleGluMetGlnValAlaGly-GlyAspValSerValIthrGly	1922
Qy	1001	AGCGGCTGGCAGAGCGCGCAGCGCTGAGCTGCAGCGCAGGAGCTTCTCATGCAGAAGC	1060
Db	1922	-----	1922
Qy	1061	GGCTGGCGATGGAGTCCACAAGATCCTCCAGGACGACGAGGATGGAGCGCANAGGA	1120
Db	1922	-----	1922
Qy	1121	GAAAAGAAATTGCCAGAAGGCAGCAGAGAAATCAGAGATACCGNAGGAGATGGAAC	1180
Db	1922	-----	1922
Qy	1181	AGATTCTAGAGGAGGAAGAGAGATTTAAGAAGCAATGGGAAGAGACTGGGGCTCAAAGG	1240
Db	1922	-----	1922
Qy	1241	AACAGCTACTTTCCTTAAACCATCTGCTGAGGTACACCCGATACCCCTTCGCGAAGC	1300
Db	1923	-----HisGlnGlnGluProAlaSerSer	1930
Qy	1301	CAAAGTGATTTCGGAAATATGAGGAAGCTTGACCCCTACTCTATGTTTCACC-----	1354
Db	1931	SerLeuSerPhe-----ThrGlyLeuThrSerSerSerIlePheGlnAspAsp	1946
Qy	1355	-----CCAGACGACATCATGGGAAGGATGTCCGGCTCTCTACGACATCAAGAGGAGGA	1408
Db	1947	LeuGlyProGlnCys-----LysSerIleThrLeuGluArg-----GlyProAsp	1962
Qy	1409	TCTTAGACCTGGCCCTGGAAGCGGTGTGNACTCCCCCATTTGG-----AAGTGGTGC	1462
Db	1963	GlyLeuGlyPheSerIleValGlyGlyTyrglyserProHisGlyAspLeuProIleTyr	1982
Qy	1463	GTTCCTGCTGTPATGCGGGGAGCTGCTGAGCGGCATGTGGCATTTGTAANGGAC	1522
Db	1983	ValIysThrValPheAlaLysGlyAlaLaserGluAspGlyArgLeuLysArgGlyAsp	2002
Qy	1523	GAGATCGCATCAACCGCAAGATTGTGACAGACTACACCTTGGCTGAGGCTGAGCCT	1582
Db	2003	GlnIlelealavalasnlglyGlnSerLeuGluGlyValThrHisGluGluAlaValAla	2022
Qy	1583	GCCTGCAGAAG	1594
Db	2023	IleLeuLysArg	2026

RESULT 10

```

RESOLD TO
US/09-922-101-3
; Sequence 3, Application US/09922101
; Patent No. US2002014671A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/922,101
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/306,998

```

```

; PRIOR FILING DATE: 1999-05-07
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-101-3

Alignment Scores:
Pred. No.: 1,79e-11 Length: 2037
Score: 263.00 Matches: 146
Percent Similarity: 36.47% Conservative: 75
Best Local Similarity: 24.09% Mismatches: 178
Query Match: 6.59% Indels: 207
DB: 10 Gaps: 26

US-09-502-945-5 (1-2162) x US-09-922-101-3 (1-2037)
Qy 11 TCGCGGTGCGGGCTCTTCCAGCTCTCGAGCCGCGGACCCGAA-----GGAACG 61
|||:||||| :||| ||| ||||| ||||| |||
Db 1552 SerGlnSerLeuMetValProGlnSerGlySerProGluProGluSerIleArgAsnThr 1571
Qy 62 GGTGCTCAACGACGAGCTGGACCTGGCCCGCCAGCCATGGACCGAAAGATGCCGCCGAGAAT 121
|||:||||| ||| ||| ||| ||| ||| ||| |||
Db 1572 SerArgSerSerThrProAlaIlePheAla-Ser--AspProAlaThrCysProIleIl 1590
Qy 122 TCGGGCATAAGGTGGATTTTCTGATTGAAATGATCG----- 158
||||| ||| ||| ||| ||| ||| ||| |||
Db 1590 eProGly-----CysGluThrThrIleGluIleSerLy 1601
Qy 159 ----AGAGAAGGACTATCTCTATGATGTCGTG----- 186
||| ||||| ||| ||| ||| ||| ||| |||
Db 1601 sGlyArgThrGlyLeuGlyLeu-SerIleValGlyGlySerAspThrLeuLeuGlyAlaI 1621
Qy 187 -----CGAATGTACCACCAG-----ACCATGGACGTGCCGTCTCTCTGTGG 226
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1621 leIleIleHisGluValThrGluGluGlyAlaAlaCysAspGlyArgLeuTrpAlaG 1641
Qy 227 GAGACCTGAAGCTGGTCATCAAT----- 249
||||| ||| ||| ||| ||| ||| ||| |||
Db 1641 lyAspGlnIleLeuGluValAsnGlyIleAspLeuArgIleAlaThrHisAspGluAlaI 1661
Qy 250 -----GAACCCAGCCGTCTGCCCTCTGTTTGATGCCATTGCGCCGCTGA 292
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1661 leAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyrArgAspGluA 1680
Qy 293 TCCCACTGAACGACACGAGTGGAAATATGATCAGCTGACCCCGCGGCTCCAGGAAGCTGA 352
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1680 laProTyrLysGluGluValCysAspThrLeuThr----- 1692
Qy 353 AGGAGTGGCTGCGGACCTCTGCCACCCGAGGCGCTCGGCTGAGTCTCGGTGGTGGCC 412
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1693 ----IleGluLeuGlnLysLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA 1711
Qy 413 TGCAGTTTGGCTCTGGGCTCTTCATCTCCACCTCATCAAAAGGGGTGAGCAGACAGACGC 472
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1711 rgAsn---AspThrGlyValPheValSerAspIleValLysGlyIleAlaAspAlaA 1730
Qy 473 TCGGG---CTCCAGTATAGGAGGAGATCGTCCGGATCAATGGATATTTCATCTCTCTCT 529
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1730 spGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGlyGluAspValArgAsnA 1750
Qy 530 GTACCCATGAGAGGTCATCAACCTCATTCGA---ACCAAGAAACTGTCTCCATCAAG 586
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1750 laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGlyThrValThrLeuGluV 1770
Qy 587 TGAGACACATCGCCCTGATCCCGTGAAGGTCF-----CCTGATGAG----- 630
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1770 alGlyArgIleLysAlaGlyProPheHisSerGluArgArgProSerGlnSerSerGlnV 1790
Qy 631 -----CCCTCCTTGGCAGTATGTGGATCATG----- 658
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


D	b	1160	SerSerThrProArgValIleProAsnValHisAsnLysAlaAsnLysIleThr	---Ser	117
Q	y	646	TATGTGATCAGATTGTGTCGGAATCTGGGGCGTCGAGGAGCGCTGGCTCCCTC	---702	
D	b	1179	AsnGlnAsnGlnAspThrGlnGluLysLysGluLysArgGlnClyThrAlaProPro	1198	
Q	y	703	-----GGAAATCGGAAACAAAGCAAG	726	
D	b	1199	MetLysLeuProProTyrLysAlaLeuThrAspAspSerAspGluAsnGluGlu	1218	
Q	y	727	AAGTCTTC-----ATCAGCTGGTAGGC-----	750	
D	b	1219	AspAlaPheThrAspGlnLysIleArgGlnArgTyrAlaAspLeuProGlyGluLeuHis	1238	
Q	y	751	-----TCCGAGGCCTGGCTGCAGCATTTCCAGCGGCCCATC	789	
D	b	1239	IleIleGluLeuGluLysAspLysAsnGlyLeuSerLeuAlaGlyAsnLysAsp	1258	
Q	y	790	CAG---AAGCCTGGCATCTTTATCAGCCATGTGAACCT---GGCTCCCTCTGCTGTGAG	843	
D	b	1259	ArgSerArgMetSerIlePheValValGlyIleAsnProGluGlyProAlaAlaAsp	1278	
Q	y	844	GTGGATGTGAGTAGGGACACAGATTGCGAAGTCAATGGCTGCAGCTTCTCTAACCTG	903	
D	b	1279	GlyArgMethIleGlyAspGluLeuGluIleAsnAsnGlnIleLeuTyrGlyArg	1298	
Q	y	904	GATCACAAGGAGCTGTAATGTGCTGTAATAATACCGCAGCTGACCATCTCCATTGTA	963	
D	b	1299	SerHisGlnAsnAlaSerAlaIleIleLysThrAlaProSer-----	1312	
Q	y	964	GCTCAGCTGGCGGAGCTGTTATGACAGACCGGGAGCGGTGCGAGGCGCGGAC	1023	
D	b	1313	-----LysValLysLeuValPheIleArgAsn	1321	
Q	y	1024	CGTGAGCTGCAGCGGAGGAGCTTCTCATGCAGAGCGGTGGCGATGAGTCCACAAG	1083	
D	b	1322	GluAspAlaValAsnGlnMetAlaValThrProPheProValProSerSerProSer	1341	
Q	y	1084	ATCTCCAGGAGCAGCAG-----GAGTGGAGCGCAAGAGGAGAGAAAGAA	1128	
D	b	1342	SerIleGluAspGlnSerGlyThrGluProIleSerSerGluGluAspGlySerLeuGlu	1361	
Q	y	1129	ATTGCCAAGAAGCCACGACGAGGAATGAGATACCGGAGGAGATGGACAGATTGTA	1188	
D	b	1362	ValGlyIleLysGlnLeuProGluSerGluSerPheLysLeuAlaValSerGlnMet---	1380	
Q	y	1189	GAGGAGGAAGAAGTTTAAAGAACAATGGGAAGAAGACTGGGCTCAAAGGAACAGCTA	1248	
D	b	1381	---LysGlnGlnLysTyr-----	1385	
Q	y	1249	CTCTTGCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTGA	1308	
D	b	1386	-----ProThrLysValSerPheSerSerGlnGluIleProLeuAlaProAlaSerSe	1403	
Q	y	1309	TTTCCGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTC-----	1351	
D	b	1403	rTyrHisSerThrAspAlaAspPheThrGlyTyrGlyGlyPheGlnAlaProLeuSerVa	1423	
Q	y	1352	-----ACCCAGACGATGTCGGGAAGGATGTCGGGTCTCATGCATCAAGAA	1401	
D	b	1423	lAspProAlaThrCysProIleValProGlyGlnGluMetIleIleGluIleSerLysGl	1443	
Q	y	1402	GGAGGGATCCTTAGACTGGCCCTCGAAGGGGTGTGGACTCCCCCATTTGGGAAGTGGT	1461	
D	b	1443	YArgSerGlyLeuSerIleValGlyLysAspThrProLeuAsnAlaIleVa	1463	
Q	y	1462	CGTTTCTGCTGTATGACGGGGAGCTGCTGAGCGGCATGGTGGCATTTGTGAAGGGGA	1521	
D	b	1463	IleHisGluValTyrGluGluGlyAlaAlaAlaArgAspGlyLeuTrpAlaGlyAs	1483	
Q	y	1522	CGAGATCATGGCAATCAACGCAAGATTGTCGACACTACACCTGCTGAGCTGACGC	1581	
D	b	1483	pGlnIleLeuGluValAsnGlnYValAspLeuArgAsnSerSerHisGluGluAlaIleTh	1503	

QY	1582	TGCGCTGCAGAGGCGCTGGAATCAGGCGGGGAGCTGGATCGACCTTGTGGTTTCCGCTCTG	1644
		:::	
Db	1503	rAlaLeuArg-----GlnThrProGlnLysValArgLeuValVal-----	1516
QY	1642	CCCCCAAGGAGTAGTATGAGCATGAGCTGACCTGCTCTTG-----	1678
		:::	
Db	1517	-----TyrArgAspGluAlaHisTyrArgAspGluGluAsnLeuGluLeuPh	1532
QY	1679	-----CTGAAGTCCAAAGGGGA-----	1696
		:	
Db	1532	eProValAspLeuGlnLysLysAlaGlyArgGlyLeuGlyLeuSerIleValGlyLysAr	1552
QY	1697	-----AACCAAATTCACGGTTAGGAACACAGTGCAGCTCCGGCC	1734
		::: :::::	
Db	1552	gAsnGlySerGlyValPheIleSerAspIleValLysGlyGlyAlaAspLeuAspG	1572
QY	1735	CCACCTCGTGAACACAAAGCCTCGGACCAACGCTTGAGAGAGCCACATCACACACCAG	1794
		: :	
Db	1572	yArgLeuIleGlnGlyAspGlnIleLeuSerValAsnGlyGluAspMetArg-----	1589
QY	1795	ATGGCATCTTGGGACCTGAATCTATCACCCAGGAATCTCAACATCCCTTTGGCCCTGAA	1854
		:::	
Db	1590	-----AsnAlaSerGlnGluThrValAlaThrIleLeuLysCysAl	1603
QY	1855	CCAGGCCAGATAGGAACAGCTCGGGCCACCTTTTGAAGGCCAATGTGGAGGAAGG	1914
		:::	
Db	1603	aGlnGlyLeuValGlnLeuGluIleGlyArgLeuArgAlaGlySer--TrpThr-----	1620
QY	1915	AGCAGCCAGCCGTTTGGGAGAGAATCTCAAGGATCCAGACTCTCATTTCTTCTCTGCG	1974
		: :	
Db	1621	SerAlaArgThrThrSerGlnAsnSerGlnGlySerGlnGlnSerAlaHisSerSer---	1639
QY	1975	CCAGTGAATTTGGTCTCTCCCAAGCTTTGGGGACACTCCTT	2013
		:::::	
Db	1640	-----CysHisProSerPheAlaProValIle	1648
RESULT 12			
US-09-997-977-3			
; Sequence 3, Application US/09997977			
; Publication No. US20030027228A1			
; GENERAL INFORMATION:			
; APPLICANT: Bartel, Paul L.			
; APPLICANT: Tavtigian, Sean V.			
; APPLICANT: Myriad Genetics, Inc.			
; TITLE OF INVENTION: MMSCL - An MMAC1 Interacting Protein			
; FILE REFERENCE: MMSCL Gene			
; CURRENT APPLICATION NUMBER: US/09/997,977			
; CURRENT FILING DATE: 2001-12-03			
; PRIOR APPLICATION NUMBER: 09/233,086			
; PRIOR FILING DATE: 1999-01-19			
; PRIOR APPLICATION NUMBER: US 60/071,861			
; PRIOR FILING DATE: 1998-01-20			
; NUMBER OF SEQ ID NOS: 65			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 3			
; LENGTH: 1881			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-09-997-977-3			
Alignment Scores:			
Pred. No.:		1.53e-10	1881
Score:		250.50	155
Percent Similarity:		36.03%	117
Best Local Similarity:		20.53%	281
Query Match:		6.28%	202
DB:		9	28
US-09-502-945-5 (1-2162) x US-09-997-977-3 (1-1881)			
Qy	70	AACGACGCGTGGACCTGGCCCGCCAGCCATGACCGCAAAAGTGGCCGAGAATTCGGCAT	129
		:::::	

Db 991 AsnAspValGlnGlyProSerLeuLeuIleAspLeuProValValAlaGlnArgArgGlu 1010
Qy 130 AAGGTGGATTTCTGATTGAAATATCATCAGAGAGGACTACTCTATGATGTGCTGCGA 189
Db 1011 GlnGluAsp-----LeuPro 1015
Qy 190 ATGTAC---CACAGACCATGACCTGCCGTCGTCGGGAGACCTGAAGCTGGTCATC 246
Db 1016 LeuTyrGlnHisGlnAla-----ThrArgValIle 1025
Qy 247 ANTGAACCCAGCGCTCTGCTGCTGTTGATGCCATTCGCCCTCATCCACTGAAGCAC 306
Db 1026 SerLysAlaSer-----AlaTyrThrGlyMetLeuSerSerArgTyr 1039
Qy 307 CAGGTGGATATGATCAGCTGACCGCCCGCGCTCCAGGAAGCTGAAGGAG----- 357
Db 1040 AlaThrAspThrCysGlnLeuProGluArgGluGluGlyGluGluThrProAsn 1059
Qy 358 -----GTGCGTCTGGACCGCTCTGCACCCCGAAGCGCTC 390
Db 1060 PheSerHisTyrGlyProProArgIleValGluIlePheArgGluProAsnValSerLeu 1079
Qy 391 GCCTGAGTGCCTGGTGGC-----CTGGAGTTTGGC----- 423
Db 1080 GlyIleSerIleValGlyGlnThrValIleLysArgLeuLysAsnGlyGluGluLeu 1099
Qy 424 TGTGGCTCTTCATCTCCACCTCATCAAGCGGTCTCAGGCAGCAGCGTC---GGGCTC 480
Db 1100 LysGlyIlePheIleLysGlnValLeuGluAspSerProAlaGlyLysThrAsnAlaLeu 1119
Qy 481 CAGGTAGGGAGCAGATCGTCGGGATCAATGGATATTCATCTCTCTGTACCCATGAG 540
Db 1120 LysThrGlyAspLysIleLeuGluValSerGlyValAspLeuGluAsnAlaSerHisSer 1139
Qy 541 GAGGTCACTCAACTCATTCGACACCAG---AAACTGTGTCTCAACTCAAGTGAACATC 597
Db 1140 GluAlaValGluAlaIleLysAsnAlaGlyAsnProValIlePheIleValGlnSerLeu 1159
Qy 598 GGC-----CTGATCCCGGTGAAGCTCTCCGTGATGAGCGCTCACTTGGCAG 645
Db 1160 SerSerThrProArgValIleProAsnValHisAsnLysAlaAsnLysIleThr---Ser 1178
Qy 646 TATGTGATCAGTTGTGCGGAATCTGGGGCGTCGAGGAGCGCTGGGCTCCGCT--- 702
Db 1179 AsnGlnAsnGlnAspThrGlnGluLysLysGluLysArgGlnGlyThrAlaProPro 1198
Qy 703 -----GGAAATCGGAAACAAAGGAGAAG 726
Db 1199 MetLysLeuProProTyrLysAlaLeuThrAspAspSerAspGluAsnGluGlu 1218
Qy 727 AAGGTCTTC-----ATCAGCTGTGTAGGC----- 750
Db 1219 AspAlaPheThrAspGlnLysIleArgGlnArgTyrAlaAspLeuProGlyGluLeuHis 1238
Qy 751 -----TCCCGAGCGCTTGGCTGCGCATTTCCAGCGGCCCATC 789
Db 1239 IleIleGluLeuGluLysAspLysAsnGlyLeuGlySerLeuAlaGlyAsnLysAsp 1258
Qy 790 CAG---AAGCTGGCATCTTATCAGCCATGTGAACCT---GGCTCCCTGTCTGCTGAG 843
Db 1259 ArgSerArgMetSerIlePheValValGlyIleAsnProGluGlyProAlaAlaAsp 1278
Qy 844 GTGGGATTTGAGATAGGGACCAAGATTTGCGAAGTCAATGGCTGCGACTTCTCTAACCTG 903
Db 1279 GlyArgMetHisIleGlyAspGluLeuLeuGluIleAsnAsnGlnIleLeuTyrGlyArg 1298
Qy 904 GATCAAGAGGCTGTAAATGTGCGAATAATAGCCGACGCTGACCATCTCCATTGTA 963
Db 1299 SerHisGlnAsnAlaSerAlaIleIleLysThrAlaProSer----- 1312
Qy 964 GCTGCAGCTGGCGGGAGCTGTTTCATCAGACAGCCGGAGCGGTGCGAGAGCGCGCAG 1023
Db 1313 -----LysValLysLeuValPheIleArgAsn 1321

Qy 1024 CGTGAGCTGCAGCGGCGAGGCTTCTCATCAGAAAGCGGTGGCGATGGAGTCCACAAG 1083
Db 1322 GluAspAlaValAsnGlnMetAlaValThrProPheProValProSerSerProSer 1341
Qy 1084 ATCTCCAGGAGCAGCAG-----CAGATGGAGCGCAAGAGAAAGAA 1128
Db 1342 SerIleGluAspGlnSerGlyThrGluProIleSerSerGluGluAspGlySerLeuGlu 1361
Qy 1129 ATTCCCAAGAGCAGCAGAGGAAATCAGAGATACCGGAGGAGATGGAAACGATTGTA 1188
Db 1362 ValGlyIleLysGlnLeuProGluSerGluSerPheLysLeuAlaValSerGlnMet--- 1380
Qy 1189 GAGGAGGAAGAGAGTTTAAGAACCAATGGGAAGACAGCTGGGCTCAAAGAAACAGCTA 1248
Db 1381 ---LysGlnGlnLysTyr----- 1385
Qy 1249 CTCTTGCCTAAACCCATCAGCTGCTGAGGTACACCCAGTACCCTTCGCAACCAAGTGA 1308
Db 1386 ---ProThrLysValSerPheSerSerGlnGluIleProLeuAlaProAlaSerSe 1403
Qy 1309 TTTCGGGAAATATGAGGAAGCGCTTGACCCCTACTCTATGTC----- 1351
Db 1403 rTyrHisSerThrAspAlaAspPheThrGlyTyrGlyGlyPheGlnAlaProLeuSerVa 1423
Qy 1352 -----ACCCAGAGCAGATCATGGGAAGGATGTCGCGCTCCTACCATCAAGAA 1401
Db 1423 LasProAlaThrCysProIleValProGlyGlnGluMetIleIleGluIleSerLysG 1443
Qy 1402 GGAGGATCCTTAGACCTGGCCCTGGGAAGCGGTGTGGACTCCCCATTTGGGAAGGTGGT 1461
Db 1443 yArgSerGlyLeuGlyLeuSerIleValGlyLysAspThrProLeuAsnAlaIleVa 1463
Qy 1462 CGTTTCTCTGTATGAGCGGAGCTGTGAGCGGATGGTGGCATTTGTGAAGGGGA 1521
Db 1463 IleHisGluValTyrGluGluGlyAlaAlaAlaArgAspGlyArgLeuTyrAlaGlyAs 1483
Qy 1522 CGAGATCATGCAATCAACGGCAAGATTGTGACAGACTACACCTGGCTGAGGTGAGCGC 1581
Db 1483 pGlnIleLeuGluValAsnGlyValAspLeuArgAsnSerSerHisGluAlaIleTh 1503
Qy 1582 TGCCCTGCGAAGCGCTGGAATCAGGCGGGAGCTGGATCGACCTGTGGTGGCTGCTG 1641
Db 1503 rAlaLeuArg-----GlnThrProGlnLysValArgLeuValVal----- 1516
Qy 1642 CCCCCAAAGAGTATGAGTACGATGACCTCTCTG----- 1678
Db 1517 -----TyrArgAspGluAlaHisTyrArgAspGluGluAsnLeuGluIlePh 1532
Qy 1679 -----CTGAAGTCCAAAGGGGA----- 1696
Db 1532 eProValAspLeuGlnLysLysAlaGlyArgGlyLeuGlyLeuSerIleValGlyLysAr 1552
Qy 1697 -----AACCAATTCACCGCTTAGGAACACAGTACAGTCCGGCC 1734
Db 1552 GAsnGlySerGlyValPheIleSerAspIleValLysGlyGlyAlaAlaAspLeuAsp 1572
Qy 1735 CCACCTCGTGAACACAAAGCGCTCGGACCGCTTGCAGAGAGGCCACATGACACACACAG 1794
Db 1572 yArgLeuIleGlnGlyAspGlnIleLeuSerValAsnGlyGluAspMetArg----- 1589
Qy 1795 ATGGCATCTTGGGACCTGAATCTATCACCAGGAATCTCAACCTCCCTTTGGCCCTGAA 1854
Db 1590 -----AsnAlaSerGlnGluThrValAlaThrIleLeuLysCysAl 1603
Qy 1855 CCAGGCCCAGATAAGAAACAGCTCGGCGCACCTTTTGAAGGCCAATGTGGAGGAAGG 1914
Db 1603 aGlnGlyLeuValGlnLeuGluIleGlyArgLeuArgAlaGlySer---TrpThr----- 1620
Qy 1915 AGCAGCCAGCCGCTTTGGGAGAGAGATCTCAAGGATCCAGACTCTCATCTTCTCTGCGC 1974
Db 1621 SerAlaArgThrThrSerGlnAsnSerGlnGlySerGlnGlnSerAlaHisSerSer--- 1639

QY 1975 CCAGTGAATTTGGTCTCTCCAGCTTTGGGGACTCCTT 2013
Db 1640 -----CysHisProSerPheAlaProValIle 1648
|||||

RESULT 13

US-09-864-761-39112
; Sequence 39112, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006684
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39112
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005137.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: SWISSPROT HIT: P31007, EVALUE 1.10e-01
; OTHER INFORMATION: EST_HUMAN HIT: BE731162.1, EVALUE 2.40e-02
US-09-864-761-39112

Alignment Scores:
Pred. No.: 3.39e-10 Length: 46
Score: 237.00 Matches: 46
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.94% Indels: 0
DB: 10 Gaps: 0

US-09-502-945-5 (1-2162) x US-09-864-761-39112 (1-46)

QY 346 AAGCTGAAGGAGGTGCTGTGACCGTCTGCACCCGAGCGCTCGGCTGAGTGTGCGT 405
Db 1 LysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeuSerValArg 20
QY 406 GTGGCCCTGGAGTTTGGCTGTGGGCTCTTCATCTCCCCACCTCATCAAAAGGGGTGAGCA 465
Db 21 GlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGlyGlyGlnAla 40
QY 466 GACACGCTCGGCTCCAG 483
Db 41 AspSerValGlyLeuGln 46

RESULT 14

US-09-919-497-59
; Sequence 59, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59
; LENGTH: 767
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-59

Alignment Scores:
Pred. No.: 1.72e-09 Length: 767
Score: 234.50 Matches: 144
Percent Similarity: 38.10% Conservative: 104
Best Local Similarity: 22.12% Mismatches: 210
Query Match: 5.88% Indels: 193
DB: 10 Gaps: 34

US-09-502-945-5 (1-2162) x US-09-919-497-59 (1-767)

QY 340 TCCAGGAGCTGAAGGAGGTGCGTCTGACCGTCTGCACCCGAGCGCTCGGCTGAGT 399
Db 198 AlaGluLysValMetGluIleLysLeuLysLeuLysLeuLysLeuLysLeuLysLeu 216
QY 400 GTGGCTGTGGTGC-----CTGGAGTTTGGCTGTGGCTGTTCATCTCCAC 444
Db 217 IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 236
QY 445 CTCATCAAGCGGTGAGGAGCGTCCAGCAGCGTCCGG---CTCCAGGTAGGGGAGAGATGCTC 501
Db 237 IleIleGluGlyGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 256
QY 502 CGGATCATGATATTCATCTCTCTGTATCCCATGAGGAGGTGATCAACCTCATTCGA 561
Db 257 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaAlaLeuLys 276
QY 562 ---ACCAAGAAACTGTGTCCATCAAGTG-----AGACACATCGGCTGATCCCC 609
Db 277 AsnThrTyrAspValValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 296

Qy	610	GTGAAAGACTCTCCTGATAGAGCCCTCACTTGG---CAGTATGTGGATCAGTTTGTCGTGC	566
Db	297	SerTyAlaProProAspPheThrThrSerTyrSerGlnHisLeuAsnGluIleSer	316
Qy	667	GAATCTGGGGCGCTGCCAGGCACCTGGGC-----	696
Db	317	Hisser-----SerTyLeuGlyThrAspTyrProThrAlaMetThrProThr	332
Qy	697	-----TCCCTCGAANTCGGAA-----AACAAAG	720
Db	333	SerProArgTyrSerProValAlaLysAspLeuLeuGlyGluGluAspIleProArg	352
Qy	721	GAGAAGAAGTCTTCATCAGCTGTGAGCTCCGAGGCCTTGCTCCACATTTCCAGC	780
Db	353	GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly	372
Qy	781	GGCCCCATCCAGAAGCTCGCATCTTTATCAGCCATG-----AAACCTGGCTCC	831
Db	373	GlyGlu---AspGlyGluGlyPheIleSerPheIleLeuAlaGlyGlyProAlaAsp	391
Qy	832	CTGTCTGCTGAGTGGGATTGGAGATAGGGACCAGATTGTCGAAGTCAATGGCTGCAC	891
Db	392	LeuSerGlyGlu-----LeuArgIysGlyAspGlnIleLeuSerValasnGlyValasp	409
Qy	892	TTCTTAACCTGGATCACAGGAGGCTGTAATGTGCTGAAAAATACCCGACCCCTGACC	951
Db	410	LeuArgAsnAlaSerHisGluGlnAlaLalleAlaLeuLysAsnAla---GlyGlnThr	428
Qy	952	ATCTCCATTCTAGCTGCAGCTGCCGGGAGCTGTTTCATGCACAGACCGGAGCGCTGGCA	1011
Db	429	ValThrIleIlealagln-----TyrLysProGluGluTyrSerArgPhe	443
Qy	1012	GAGCGCGGCGAGCTGAGCTGCAGCGGCAGGAGCTTCTCATCGAAGCGGCTGGCGATG	1071
Db	444	GluAlatylsileHisAspLeuArgGluGln----LeuMetAsnSerSerLeuGlySer	461
Qy	1072	GAGTCCAACAAGATCTCTCAGGNAGCAGCAGAG-----ATGCAG	1110
Db	462	GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrlleArgAlaLeuPheAsp	481
Qy	1111	CGGCAAGGAGAAAGAAAATTGCCAGAGGCAGCAGAGGAAAATGAGATATCCGGAAG	1170
Db	482	TyrAspLysThrLysAspcysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp	501
Qy	1171	GAGATGGAACAGATTCTAGAGGAGGAAGAGAAGTTT-----	1206
Db	502	ValLeuHisValIleAspAlaSerAspGluGluTrpTrpGlnAlaArgArgValHissSer	521
Qy	1207	-----AAGAAGCAA	1215
Db	522	AspSerGluThrAspAspIleGlyPheIleProSerLysArgArgValIGluArgArgGlu	541
Qy	1216	TGGCAA-----GAAGACTCGGGCTCA-----AAGGAA	1242
Db	542	TrpSerArgLeuLysAlaLysAspTrpGlySerSerGlySerGlnGlyArgGluAsp	561
Qy	1243	CAGTACTCTTGCTATAAACCACTACFT---GCTGGTAGTACACCCAGTACCCCTTCGCAAG	1299
Db	562	SerValLeuSerTyrGluThrValThrGlnMetGluValHisTyrAlaArg-----	578
Qy	1300	CAAAGTGATTTCGGGAAATATAGGAAGGCTTTGACCCCTACTCTATGTTCACCCCAGA	1359
Db	579	-----ProIleIleLeuGlyProThr	586
Qy	1360	CGCATCATGGGNAGGATGTCGGGCTCCTACGCATCAAGAAGGAGGATCCTTAGACCT	1419
Db	587	LysAspArgala-----AsnAspAspLeuLeuSerGlu	597
Qy	1420	GGCCCTGGAGGGCGGTGTGGACTCCCCCATTTGGGAAGGTGGTCTTCTGCTGCTATGA	1479
Db	598	PheProAspyys-----	601
Qy	1480	CGGGGAGCTCTGAGCGGCATGCTGGCATTTGTGAAAGG--GGACGAGATCATGGCAAT	1536

```

; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39113
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005137.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EST_HUMAN HIT: BE266204.1, EVALUOE 5.30e+00
US-09-864-761-39113

```

```

Alignment Scores:
Pred. No.:      8.05e-10      Length:      45
Score:          232.00      Matches:      39
Percent Similarity: 88.84%      Conservative: 0
Best Local Similarity: 88.64%      Mismatches: 5
Query Match:      5.82%      Indels:      0
DB:              10      Gaps:      0

```

US-09-502-945-5 (1-2162) x US-09-864-761-39113 (1-45)

```

Qy 321 TCAGCTGACCCCGCGCCTCCAGGAGCTGAAGAGGTGCGTCTGACCGTCTGACCC 380
Db 2 SerProThrSerProValTrpGlnGluAlaGluGlyAlaSerGlyProSerAlaPro 21
Qy 381 CGAAGGCGCTCGCCTGAGTGTGCTGGTGGCTGGAGTTGGCTGGGCTCTTCATCTC 440
Db 22 ArgArgProArgProGluCysAlaTrpTrpProGlyValTrpLeuTrpAlaLeuHisLeu 41
Qy 441 CCACCTCATCAA 452
Db 42 ProProHisGln 45

```

Search completed: March 21, 2003, 13:58:21
Job time : 52.1498 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:32:38 ; Search time 38.5119 Seconds
(without alignments)
10793.685 Million cell updates/sec

Title: US-09-502-945-5
Perfect score: 3989
Sequence: 1 cctggccggtcggtgcg.....tattttccagcttaaaaaa 2162

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlp
-O=/cgn2_1/USPRO_pool/US09502945/runat_14032003_101059_19127/app_query.fasta_1.10979
-DB=PIR-73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DICALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_283 -runat_14032003_101059_19127 -NCPU=6 -ICPU=3
-NO_XLPYX -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR-73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	269.5	6.8	2054	2 T46612	multi PDZ domain p
2	259.5	6.5	2055	2 T30259	multiple PDZ domai
3	244.5	6.1	720	2 A45436	synapse-associated
4	243	6.1	1012	2 T23160	hypothetical prote
5	236.5	5.9	926	2 I38756	homolog of Drosoph
6	236	5.9	1745	2 A46431	tight junction-ass
7	234.5	5.9	767	2 T09599	postsynaptic densi
8	232.5	5.8	724	2 JH0800	postsynaptic densi
9	232	5.8	904	2 I38757	homolog of Drosoph
10	227	5.7	2450	2 S71625	protein-tyrosine-p
11	226	5.7	911	2 I56552	synapse-associated
12	226	5.7	1131	2 T15617	hypothetical prote
13	225	5.6	1736	2 A47747	tight junction pro
14	218.5	5.5	852	2 T10811	channel associated

15	213.5	5.4	870	2 G01974	channel associated
16	211	5.3	1112	2 T32733	AMPA glutamate rec
17	206	5.2	723	2 T14765	hypothetical prote
18	204.5	5.1	2466	2 T16729	protein-tyrosine-p
19	204	5.1	1367	2 T13703	tama protein - fru
20	203.5	5.1	628	2 T09458	numb-binding prote
21	203.5	5.1	728	2 T09457	numb-binding prote
22	203	5.1	2294	2 T16730	protein-tyrosine-p
23	202.5	5.1	1277	2 T14152	synaptic scaffold
24	194.5	4.9	450	2 G01158	tyrosine kinase ac
25	193.5	5.0	2715	2 T13049	eyelid - fruit fly
26	192	4.8	2490	1 A54971	protein-tyrosine-p
27	191.5	4.8	1256	2 JEO209	brain-specific ang
28	190	4.8	1337	2 T13948	atypical protein k
29	189.5	4.8	2172	2 T20145	hypothetical prote
30	188.5	4.7	358	2 I46532	protein co-factor
31	188	4.7	1464	2 T13716	bazooka gene prote
32	185	4.6	1116	2 I54378	gene X104 protein
33	183	4.6	1281	2 T00346	hypothetical prote
34	181.5	4.6	1163	2 JEO366	tight junction pro
35	179	4.6	660	1 Q0BE3	BHLFI protein - hu
36	179	4.5	1171	2 T42372	probable guanylate
37	176.5	4.5	1006	2 T42731	atrophin-1 related
38	171	4.3	817	2 T03852	protein phosphatas
39	171	4.3	1553	2 T03301	rab3 effector prot
40	167.5	4.2	1373	1 A43291	collagen alpha 2(I
41	166	4.2	960	1 A39651	discs-large tumor
42	164.5	4.1	1414	1 S23809	collagen alpha 2(I
43	164	4.1	640	2 T08179	LRG5 protein - chl
44	164	4.1	1453	2 S21626	collagen alpha 1(I
45	164	4.1	1464	1 CGH01S	collagen alpha 1(I

ALIGNMENTS

RESULT 1

T46612

multi PDZ domain protein 1 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 21-Jul-2000

C:Accession: T46612

R:Ullmer, C.; Schmuck, K.; Figge, A.; Lubbert, H.

FEBS Lett. 424, 63-68, 1998

A:Title: Cloning and characterization of MUPP1, a novel PDZ domain protein.

A:Reference number: 223104; MUID:98196865; PMID:9537516

A:Accession: T46612

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-2054 <ULL>

A:Cross-references: EMBL:AJ001320; NID:g2959978; PIDN:CAA04681.1; PID:g2959979

A:Experimental source: brain

C:Genetics:

A:Gene: MUPP1

Alignment Scores:			
Pred. No.:	8.2e-10	Length:	2054
Score:	269.50	Matches:	120
Percent Similarity:	38.20%	Conservative:	71
Best Local Similarity:	24.00%	Mismatches:	161
Query Match:	6.76%	Indels:	148
DB:	2	Gaps:	17

US-09-502-945-5 (1-2162) x T46612 (1-2054)

QY	157	GCAGAGAAGCATCTCTATGATGTCGCAATGTACCAACACGACGACGTGGCC	216
Db	1672	AlaThrHisAspGluAlaIleAsnValLeuArgInThrProGlnArgVal	1688
QY	217	GTGCTCGTGGGAGACCTGAAGCTGGTCAATCAATGAACCCAGCCGCTGTGTTTGTAT	276
Db	1689	-----ArgLeuThrLeuTyr----	1693
QY	277	GCAATTCGGCCGCTGATCCCACTGAAGCACCAGTGAATATGATCAGTGCACCCCGG	336


```
Db 1694 -----ArgAspGluAlaProTyrLysGluGluAspValCysAspThrPheThr----- 1709
QY 337 CGCTCCAGGAGCTGAAGGAGTGGCTGTGACCGTCTGCACCCCGAGGCGCTGGCGCTG 396
Db 1710 -----ValGluLeuGlnLysArgProGlyLysGlyLeuGlyLeu 1722
QY 397 AGTGTGGCTGGTGGCTGGAGTTGGTGTGGGCTCTTCATCTCCACCTCATCAAAAGGC 456
Db 1723 SerileValGlyLysArgAsn---AspThrGlyValPheValSerAspIleValLysGly 1741
QY 457 GGTGAGGAGACAGCGTGGG---CTCAGTAGGAGGAGATCGTCCGGATCAATGGA 513
Db 1742 GlyIleAlaAspAlaAspGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGly 1761
QY 514 TATTCATCTCTCTGTACCATGAGGAGTGCATCAACCTCATTCGA---ACCAAGAAA 570
Db 1762 GluAspValArgAsnAlaThrGlnGluAlaValAlaAlaLeuLysCysSerLeuGly 1781
QY 571 ACTGTGTCCATCAAGTGAGACATCGCTGATCCCGTGAAAGCTCTCTCGATGAG 630
Db 1782 ThrValThrLeuGluValGlyArgIleLysAlaAlaProPheHisSerGluArgArgPro 1801
QY 631 CCCCTCACTGGCAGTATGTGATCAGTTGTGTGCGAATCTGGGGCGCTCGAGGCAGC 690
Db 1802 SerGlnSerSerGlnValSerSerLeuSerPheSerLeuProArgSerGly 1821
QY 691 CTGGGCTCCCTGGAATCGGGAACAGGAGAGAAG----- 729
Db 1822 IleHisThrSerGluSerSerAlaLysLysAlaLeuAlaSerGluIle 1841
QY 730 -----GTCTTCATCAGCTGCTGAGTCCCGAGGCTTGGCTGCAACATT 774
Db 1842 GlnGlyLeuArgThrValGluIleLysLysGlyProAlaAspAlaLeuGlyLeuSerIle 1861
QY 775 TCCAGCGG-----CCCATCCAGAGCGCTGCATCTTATCAGCCATGTGAACCT 825
Db 1862 AlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetHisPro 1881
QY 826 GGCTCCCTGCTGCTGAGGTG---GGATTGGAGATAGGGGACAGATGTCAAGTCAAT 882
Db 1882 AsnGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleCys 1901
QY 883 GCGGTGCACTTCTTAACCTGATCACAAGAGGCTGTAATGTGCTGAAATAGCCGC 942
Db 1902 GlyThrSerThrAspGlyMetThrHisThrGlnAlaValAsnLeuMetLysAsnAla--- 1920
QY 943 AGCCTGACCATCTCCATCTGATCGCTGCGCGGAGCTGTTTCATGACAGACCGGAG 1002
Db 1921 SerGlySerIleGluValGlnValValAlaGlyAspVal----- 1934
QY 1003 CGGCTGGCAGGCGCGGCGGCTGAGCTGCAGCGGAGGAGCTCTCATGAGAGCGG 1062
Db 1935 SerValValThrGlyHisGlnGlnGluLeuAlaAsnProCysLeuAlaPhe----- 1951
QY 1063 CTGGCGATGAGTCCCAACAAAGATCTCCAGGAGCAGCAGGAGATGGAGCGCAAGGAGA 1122
Db 1952 ThrGlyLeuThrSerSerThrIle----- 1959
QY 1123 AAGAAATTTGCCAGAGCAGCAGAGGAAATGAGAGATACCGGAGGAGATGGAACAG 1182
Db 1959 ----- 1959
QY 1183 ATTGTAGAGGAGAGAGAAATTTAAGAACATGGGAAGAGACTGGGCTCAAGGAA 1242
Db 1960 -----PheProAspAspLeuGlyProProGln 1968
QY 1243 CAGCTACTCTGCTAAACCACTACTGCTGAGGTACACCCAGTACCCCTCGCAAGCCA 1302
Db 1969 Ser-----LysThrIleThr----- 1973
QY 1303 AGTGTATTTCCGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTTACCCCGAGCA 1362
```

```
Db 1974 -----Leu 1974
QY 1363 GATCATGGGAAGAGATGTCGGCTCTACGCATCAAGAAGGAGGATCCTTAGACCTGGC 1422
Db 1975 AspArg-GlyProAsp-----GlyLeuGlyPhe 1984
QY 1423 CTGGAAGGCGGTGTGAGTCCCTCCCATTTGG-----AAGGTGCTGCTTCTCTCTGTA 1476
Db 1984 rIleValGlyGlyTyrGlySerProHisGlyAspLeuProIleTyrValLysThrValPhe 2004
QY 1477 TGAGCGGGGAGCTGCTGAGCGGCATGCTGTCATTGTGAAGGGGAGAGATCATGGCAAT 1536
Db 2004 eAlaLysGlyAlaAlaAlaGluAspGlyArgLeuLysArgGlyAspGlnIleAlaVal 2024
QY 1537 CAACGGCAAGATTGTACAGACTACACCTGGCTGGCTGAGGTGAGCTGCGCTCCCTGCAGAG 1594
Db 2024 lAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaValAlaIleLeuLysArg 2043
RESULT 2
T30259
multiple PDZ domain protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30259
R:Simpson, E.H.; Suffolk, R.; Jackson, I.J.
Genomics 59, 102-104, 1999
A:Title: Identification, sequence, and mapping of mouse multiple PDZ domain protein 9
A:Reference number: Z20797; MUID:99326529; PMID:10395806
A:Accession: T30259
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2055 <SIM>
A:Cross-references: EMBL:AJ131869; NID:94150877; PIDN:CAA10523.1; PID:94150878
A:Experimental source: strain C57/BL6 x CBA Fl; whole brain
C:Genetics:
A:Gene: mpdz
A:Map position: 4
Alignment Scores:
Pred. No.: 3,82e-09 Length: 2055
Score: 259.50 Matches: 117
Percent Similarity: 37.45% Conservative: 71
Best Local Similarity: 23.31% Mismatches: 162
Query Match: 6.51% Indels: 152
DB: 2 Gaps: 19
US-09-502-945-5 (1-2162) x T30259 (1-2055)
QY 157 GCAGAGAAGGACTATCTATGATGCTGCGAATGTACACACAGACCATGGACGTGGCC 216
Db 1673 AlaThrHisAspGluAlaIleAsnValLeuArgGlnThrProGlnArgVal----- 1689
QY 217 GTGCTGTGGGACAGCTGAAGCTGTGTCATCAATGAACCCAGCCGCTCTCTCTTTGAT 276
Db 1690 -----ArgLeuThrLeuTyr--- 1694
QY 277 GCCATTGCGCGCTGATCCCACTGAAGCACCAAGTGGATATGATCAGCTGACCCCGGG 336
Db 1695 -----ArgAspGluAlaProTyrLysGluGluAspValCysAspThrPheThr----- 1710
QY 337 CGCTCCAGGAGCTGAAGGAGTGGCTGACCGCTGACCGCTGACCCCGAGGCGCTCGGCTG 396
Db 1711 -----IleGluLeuGlnLysArgProGlyLysGlyLeuGlyLeu 1723
QY 397 AGTGTGCTGCTGGCGCTGGAGTTGGCTGTGGCTTCTTCATCTCCACCTCATCAAGGC 456
Db 1724 SerIleValGlyLysArgAsn---AspThrGlyValPheValSerAspIleValLysGly 1742
QY 457 GGTGAGGAGACAGCGTGGG---CTCAGTAGGAGGAGATCGTCCGGATCAATGGA 513
Db 1743 GlyIleAlaAspAlaAspGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGly 1762
QY 514 TATTCATCTCTCTGTAACCATGAGGAGTGTACCCATGAGGAGTCAATCAACCTCATCGA---ACCAGAAA 570
```

[illegible]

```
Db   2023 eAlaValAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaValAlaIleLeuLy 2043
Qy   1591 GAAG 1594
Db   2043 sarg 2044
RESULT 3
A45436
synapse-associated protein SAP90 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000
C:Accession: A45436
R:Kistner, U.; Wenzel, B.M.; Veh, R.W.; Cases-Langhoff, C.; Garner, A.M.; Appeltauer,
J. Biol. Chem. 268, 4580-4583, 1993
A:Title: SAP90, a rat presynaptic protein related to the product of the Drosophila tu
A:Reference number: A45436; MUID:93186749; PMID:7680343
A:Accession: A45436
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-720 <KIS>
A:Experimental source: brain
A>Note: sequence extracted from NCBI backbone (NCBIP:126555)
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
F:70-148/Domain: GLGF domain homology <GLGF>
F:165-243/Domain: GLGF domain homology <GLGF>
F:435-493/Domain: SH3 homology <SH3>
F:531-708/Domain: guanylate kinase homology <GKI>

Alignment Scores:
Pred. No.:      2,61e-08      Length:      720
Score:          244.50        Matches:    144
Percent Similarity: 38.18%     Conservative: 103
Best Local Similarity: 22.26%   Mismatches:  211
Query Match:     6.13%         Indels:     189
DB:              2           Gaps:         33

US-09-502-945-5 (1-2162) x A45436 (1-720) .
Qy   340 TCCAGGAAGCTGTAAGGAGGTGGCTGTGCACGCCCGAACGGCCGTCGCAGT 399
Db   155 AlaGluysValMetGluiLleLysLeullefys---GlyProlysGlyLeuGlyPheSer 173
Qy   400 GTGGCGTGTTGGC-----CTGAGATTGGCTGTGGCTGCTCATCTCCCAC 444
Db   174 IlealagGlylvalGlyAsnGlnHisfileproGlyAspAsnSerileTyValThrLys 193
Qy   445 CTCATCAAGCGGTGACGAGACAGCGTCGGG---CTCACGATAGGGGACGAGATCGTC 501
Db   194 llelleGluGlylAlaAlaHisLysAspGlyArgLeuGlnleGlyAsplylleLeu 213
Qy   502 CGGATCAATGATATTCCATCTCTCTGTACCATGAGGAGTCATCAACCTCATTCGA 561
Db   214 AlalaAsnSerValGlyLeuGluaspValmethHisgluaspaAlaAlaLeuLys 233
Qy   562 ---ACCAAGAAACTGTGTCATCAAAGTG-----AGACACATCGGCCTGATCCCC 609
Db   234 AsnthrtyraspvalvalTyLeuLysvalalalySproSerAsnAlatyLeuSerAsp 253
Qy   610 GTGAAAAGCTCTCTGTATGAGCCCCCTCAGTGG---CAGTATGTGGATCAGTTGTGTGC 666
Db   254 SerTyralaProProaspilleThrThrSertyserGlnHisLeuaspAsnGluiLeSer 273
Qy   667 GAATCTGGGGCGTGGCAGGACGCTGGGC----- 696
Db   274 HisSer-----SertyrLeuGlylThrAsptyrProThralamethThrProThr 289
Qy   697 -----TCCCTGGAAATCGGGAA-----AAACAAG 720
Db   290 SerProArgArgTySerProValAlaLysaspleuLeuGlyGluGluaspilleProArg 309
Qy   721 GAGAAGAAGGTCATCATCAGCGTGGTAGGCTCCCGAGGCCTTCGGCTGCACGATTCACGC 780
Db   310 GluproArqarlleValilleHsisArgGlySerThrglyLeuglyPheAsnilleValGlv 329
```

```
QY 781 GGCCTCCAGAGCCCTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
  |||
  |||
Db 330 GlyGlu---AspGlyGluGlyPheIleLeuAlaGlyGlyProAlaAsp 348
  |||
  |||
QY 832 CTGCTCTGCTGAGGTGGGATTTGGAGATAGGGACAGATGTCGAAGTCAATGGCGCTCAG 891
  |||
  |||
Db 349 LeuSerGlyGlu-----LeuArgGlyGlyAspGlnIleLeuSerValAsnGlyValAsp 366
  |||
  |||
QY 892 TTTCTTAACCTGATCAAGAGGCTGTAAATGTGCTGAAATATAGCGCGACCTGACC 951
  |||
  |||
Db 367 LeuArgAsnAlaSerHisGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 385
  |||
  |||
QY 952 ATCTCATTTAGTCTGACGCTGGCGGAGCTGTTTCATCAGACACCGGAGCGCTGGCA 1011
  |||
  |||
Db 386 ValThrIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
  |||
  |||
QY 1012 GAGGCGGCGCAGCTGAGCTGAGCGGCGGAGGAGCTTCTCATCCAGAGCGGCTGCGATG 1071
  |||
  |||
Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
  |||
  |||
QY 1072 GAGTCCAAACAAGATCTCCAGGAGCAGCAGGAG-----ATGGAG 1110
  |||
  |||
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
  |||
  |||
QY 1111 CGSCAAAGGAGAAAGAAATTGCCAAGAGCAGCAGAGGAAATCAGAGATACCGGAAG 1170
  |||
  |||
Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 458
  |||
  |||
QY 1171 GAGATGGACAGATCTGATAGAGGAGGAGGAGAGATT----- 1206
  |||
  |||
Db 459 ValLeuHisValIleAspAlaGlyAspGluGluTrpTrpGlnAlaArgValHisSer 478
  |||
  |||
QY 1207 -----AAGAAGCAA 1215
  |||
  |||
Db 479 AspSerGluThrAspAspIleGlyPheIleProSerLysArgArgValGluArgGlu 498
  |||
  |||
QY 1216 TGGGAAGAAGACTGGGGCTCA-----AAGGAACGACTACTCTTG 1254
  |||
  |||
Db 499 TrpSerArgLeuTrpGlySerSerGlySerGlnGlyArgGluAspSerValLeuSer 518
  |||
  |||
QY 1255 CCTAAACCATCACT---GCTGAGGTACACCCAGTACCCCTTCGCGAAGCAAGTGATT 1311
  |||
  |||
Db 519 TyrGluThrValThrGlnMetGluValHisTyrAlaArg----- 531
  |||
  |||
QY 1312 CCGGAATATGAGGAGGCTTTGACCCCTACTCTATGTTCCACCCAGCAGCATGGG 1371
  |||
  |||
Db 532 -----ProIleIleLeuGlyProThrLysAspArgAla 543
  |||
  |||
QY 1372 GAAGGATGTCCGGCTCTCTAGCATCAAGAAGGAGGATCTTAGACCTGGCCCTGGAAGG 1431
  |||
  |||
Db 544 -----AsnAspAspLeuLeuSerProSerProAspLys 554
  |||
  |||
QY 1432 CGGTGTGGACTCCCATTTGGGAAGTGGTCTGCTGTATGACGGGGGAGCTGC 1491
  |||
  |||
Db 555 -----PheGlySerCys 558
  |||
  |||
QY 1492 TGAGCGCATGTGGCATTGTGAAGG---GGACGAGATCATGGCAATCAACGGCAAGAT 1548
  |||
  |||
Db 559 ValPro-----HisThrThrArgProLysArgGluTyr---GluIleAspGlyArg-- 574
  |||
  |||
QY 1549 TGTGACAGACTACCCCTGGCTGAGGCTGACGCTGCCCTGCAGAGGCGCTGGGAATCAGGG 1608
  |||
  |||
Db 575 -----AspTyrHisPheValSerArgGluLysMetGluLys----- 587
  |||
  |||
QY 1609 CGGGGACTGATCGACCTTGTGTTGCCCTCTGCCCCCAAGAGGATGATGACCATGAGCT 1668
  |||
  |||
Db 588 -----AspIleGln 591
  |||
  |||
QY 1669 GACCTTCTTCTGAAGTCCAAAGGGGAAACCAATTCACCGCTTAGGAACAGT----- 1723
  |||
  |||
Db 591 aHisLysPheIleGluAlaGlyGlnTyrAsnSer---HisLeuTyrGlyThrSerValGln 610
  |||
  |||
```

```
QY 1724 -----GAGCTCCGGCCCGCCACCTCGTGAACACAAAGCCTCGGACCAG 1764
  |||
  |||
Db 610 nserValargGluValAlaGluGlnGlyLysHisCysIleLeuAspValSerAlaAsnAl 630
  |||
  |||
QY 1765 CTTGAGAGAGGCGCACATGACACACACAGATGCATCTTGGGACCTGAATCTATCACC 1824
  |||
  |||
Db 630 aValArgLeuGlnAlaAlaHisLeu-----HisProIleAlaIlePheIleArgPr 648
  |||
  |||
QY 1825 CAGGAATCTCAAACTCCCTTTGGCCCTGAAC-----CAGGCCAGAT 1866
  |||
  |||
Db 648 oArgSerLeuGluAsnValLeuGluIleAsnLysArgIleThrGluGlnAlaArgL 668
  |||
  |||
QY 1867 AAGGAACAGCTCGGCGCACCTTTTGTGAAG-----GCCAATG 1902
  |||
  |||
Db 668 ysAlaPheAspArgAlaThrLysLeuGluGlnGluPheThrGluCysPheSerAlaIleV 688
  |||
  |||
QY 1903 TGGGAGAAAGGAGCAGCAGCCGCTTTGGGAGAAGATCTCAAGG---ATCCAGACTCTCA 1959
  |||
  |||
Db 688 alGluGlyAspSerPheGluGluIleTyrHisLysValLysArgValIleGluAspLeuS 708
  |||
  |||
QY 1960 TT---CCTTTCTCTCTGG 1973
  |||
  |||
Db 708 erGlyProTyrIleTrp 713
  |||
  |||
RESULT 4
T23160
Hypothetical protein K01A6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C:Accession: T23160
R:Cottage, A.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19701
A:Accession: T23160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1012 <WIL>
A:Cross-references: EMBL:Z68750; PIDN:CAA92965.1; GSPDB:GN00022; CESP:K01A6.1
A:Experimental source: clone K01A6
C:Genetics:
A:Gene: CESP:K01A6.1
A:Map position: 4
A:Introns: 107/1; 151/2; 183/2; 270/2; 316/3; 410/3; 628/3; 922/3; 971/3
A:Superfamily: WW repeat homology
F:131-168/Domain: WW repeat homology <WWR>

Alignment Scores:
Pred. No.: 3,73e-08 Length: 1012
Score: 243.00 Matches: 123
Percent Similarity: 36.47% Conservative: 71
Best Local Similarity: 23.12% Mismatches: 191
Query Match: 6.09% Indels: 147
DB: 2 Gaps: 19

US-09-502-945-5 (1-2162) x T23160 (1-1012)
QY 29 CCAGCTCTCGCAGCGGGGCAACCCGCAAGAGCGGTCTG-----GCACGACG 76
  |||
  |||
Db 574 ProLysThrArgSerArgThrProSerAlaAlaPheArgTyrGlyGluProGlnThrAsn 593
  |||
  |||
QY 77 CAGCTGGACCTGGCCGACGCA-----TGGACC 103
  |||
  |||
Db 594 MetMetAspSerAlaAlaProLeuProValArgSerLysThrProAlaGluArgGlnThr 613
  |||
  |||
QY 104 GAAAGTGGCCCGAGAAATTCGCGCATAGGTGGATTTCTGATTGAAATGATGCAGAGA 163
  |||
  |||
Db 614 SerArgThrGluGluAspGlnAsnValArg-----AsnThrLeuGlnArg 628
  |||
  |||
QY 164 AGGACTATCTCTATGATGTGCTGCGAATGTACCAACGACCATGGACGCTGGCGCTCG 223
  |||
  |||
Db 629 Gln-----ProAlaValThrSerGluTrp---GluGlyMetSerSe 641
  |||
  |||
QY 224 TGGGAGACCTGAAGCTGGTCAATCAATGAACCCAGCCGCTGCTGCTCTGTTTGTGCAATTC 283
  |||
  |||
```

```
Db 641 rAlaIlePro--AlaSerArgMet-ArgProSerSerThrThrLeuGlyPheAlaThrP 660
QY 284 GCCCGCTATCCACTGAAG-----CACAGGTGGATATATGATCAGCTGACCCCGCCG 337
Db 660 roAsnTyrIleProLeuSerGlnTyrAsnGlnLysProSerAspLeuIleThrValSerL 680
QY 338 GCTCAGGAAGCTGAAGAGGTGCTCTGGACCGTCTGCACCGAGGCGCTCGGCCTGA 397
Db 680 euIleArgLys-----ProValGlyPheGlyPheA 690
QY 398 GTGTCGGTGGCTGGAGTTTGGCTGTGGCTCTTCCATCCACCTCATCAAGGCG 457
Db 690 rgLeuLeuGlyValGlySerLysThrProLeuSerValGlyGlnIleValIleGlyG 710
QY 458 GTCAGGCACAGCGTCGG---CTCCAGGTAGGGAGCAGATCGTCGGGATCAATGAT 514
Db 710 lYAlaAlaGluGluAspGlyArgLeuGluGlyAspGlyIleValGluIleAspGlyH 730
QY 515 ATTCCATCTCTCTGTACCATGAGGAGTCACTCAACCTCATTCGACCAAGAAACTG 574
Db 730 iAsnValGluGlyAlaSerHisSerGluAlaValValLeuLeuGlu-----A 746
QY 575 TGTCATCAAGTGAACATCGCCCTGATCCCGTGAAGAGCTCTCTCATGAGCCCC 634
Db 746 laAlaAlaGlnAsnLysHisValLysLeuIleValArgProSerArgThrAspProA 766
QY 635 TCACCTGGCAGTATGTGATCAGTTTGTGCGGAATCTGGGGCGTGGAGCGACCTGG 694
Db 766 laArg-----ArgGlySerLeuA 772
QY 695 GCTCCCTCGGAATCGGAAACAGAGAGAGTCTTTCATCAGCCTGTAGGCTCCC 754
Db 772 snSerAlaGlyProSerGlySerThrAsp-----ValLeuLeuHisArgAsnGluAsnA 790
QY 755 GAGCGCTTGGCTGAGCAGATTTCCAGCGCCCATCCAGAACCTGGCATCTTTATCAGCC 814
Db 790 spGlyPheGlyPheValLeuMetSerSerGln---HisLysAsnGlySerThrValGlyG 809
QY 815 ATGTGAACCTGGCTCCCTGCTCTGAGTGGGA---TTGAGATAGGGAGCAGATTG 871
Db 809 lnIleGlnProGlySerProAlaAlaArgCysGlyArgLeuSerValGlyAspArgValI 829
QY 872 TCGAGTCAATGGGTGCAGCTTCTTAACCTGGATCAACAGGAGGCTTAATGTGTGA 931
Db 829 leAlaValAsnGlyLysPheValLeuSerLeuSerHisProAspThrIleSerLeuIleL 849
QY 932 AAATAGCCGC-----AGCTGACCATCTCCATTTAGCTGCAGCTGCGCGGAGC 982
Db 849 ysAspSerGlyLeuSerValArgLeuThrIleAlaProProAsnThrAlaGlyProValL 869
QY 983 TGTTCATCAGACCGGAGCGGTGGCGAGTGGAGTCCCAACAGATCTCCAGGAGCAGCAGG 1042
Db 869 euProMetVal-SerAlaThr-----LeuGlyArg 878
QY 1043 AGCTTCTCATCAGAAGCGGTGGCGATGGAGTCCCAACAGATCTCCAGGAGCAGCAGG 1102
Db 879 AsnPheThr----- 881
QY 1103 AGATGGAGCGCAAGGAGAAAGAAATTTGCCAGAGGCAGCAGAGGAAATGAGAGAT 1162
Db 881 ----- 881
QY 1163 ACCGGAAGGAGATGGAACAGATTCTAGAGGAGGAGAGAGAGATTTAAGAACGAATGGGAAG 1222
Db 881 ----- 881
QY 1223 AAGACTGGGGCTAAAGAACAGCTACTTTGGCTTAACCATCACTGCTCAGGTACACC 1282
Db 882 MetAsnGlyHisTyrGluSerAsnTyrGlyLeuProProProProSerValTyrGlu 901
QY 1283 CAGTACCCCTTCGCAAGCCAAAGTGATTTCGGGAATATGAGGAAGCTTTGACCCCTAC 1342
Db 901 ----- 1342
```

```
Db 902 LysHisProProProSerTyrLeu-----AlaPheAspGlyLeu 914
QY 1343 TCTATGTTCACCCAGACAGCATCATGGGAAG-----GATGTCCGGCTCTTACGCATC 1396
Db 915 SerIleAsnAspArgMetSerMetAsnGlyAsnLeuIleAspValThrLeuGluArgGly 934
QY 1397 AAGAAGAGGAGGATCTTAGACCTGGCCCTGGAAGGCGGTGGACTCCCCCATTTGGGAAG 1456
Db 935 ThrLys-----GlyPheGlyPheSerIleArgGlyGlnGluPheGlySerMetPro 952
QY 1457 GTGTCCTTCTTCTGTGTATGAGCGGGAGCTGCTCAGCGGATGGTCATTTGTGAAA 1516
Db 953 LeuPheValLeuArgIleAlaAspAspGlyProAlaLysAlaAspGlyArgLeuGlnVal 972
QY 1517 GGGACAGAGTCAATGCAATCAACGCGCAAG 1546
Db 973 GlyAspGlnLeuThrThrIleAsnGlyGln 982
RESULT 5
138756
homolog of Drosophila discs large protein, isoform 2 - human
C:Species: Homo sapiens (man)
C>Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
C:Accession: I38756
R:Lue, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.
Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila
A:Reference number: I38756; MUID:95024052; PMID:7937897
A:Accession: I38756
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-926 <RES>
A:Cross-references: EMBL:U13896; NID:g558435; PIDN:AAA50598.1; PID:g558436
C:Genetics:
A:Gene: DDLG1
A:Cross-references: GDB:393278; OMIM:601014
A:Map position: 3q29-3q29
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
C:Keywords: alternative splicing; duplication
F:229-307/Domain: GLGF domain homology <GLG1>
F:324-402/Domain: GLGF domain homology <GLG2>
F:588-646/Domain: SH3 homology <SH3>
F:737-914/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 9.8e-08 Length: 926
Score: 236.50 Matches: 133
Percent Similarity: 40.07% Conservative: 97
Best Local Similarity: 23.17% Mismatches: 214
Query Match: 5.93% Indels: 130
DB: 2 Gaps: 27
US-09-502-945-5 (1-2162) x I38756 (1-926)
QY 340 TCAGGAAGCTGAAGAGGTGCGTGTGACCGCTGCACCCGAGGCGCTCGGCCTGAGT 399
Db 314 SerGluLysIleMetGluIleLysLeuIleLys---GlyProLysGlyLeuGlyPheSer 332
QY 400 GTGCGTGGTGGC-----CTGGAGTTTGGCTGTGGGCTCTTCATCTCCAC 444
Db 333 IleAlaGlyValGlyAlaGlnHisIleProGlyAspAsnSerIleTyrValThrLys 352
QY 445 CTCATCAAGCGCGTCAGGCAGACAGCGTCGGG---CTCCAGTAGGGACGAGATCGTC 501
Db 353 IleIleGluGlyGlyAlaAlaHisLysAspGlyLysLeuIleGlyAspLysLeuLeu 372
QY 502 CGGATCAATCGATATTCATCTCTCTCTTACCATCAGGAGGTCATCACTTCGA 561
Db 373 AlaValAsnAsnValCysLeuGluGluValThrHisGluGluAlaValThrAlaLeuLys 392
QY 562 ---ACCAAGAAAACCTGTGTCTCATCAAAAGTG-----AGACACATC 597
Db 393 AsnThrSerAspPheValTyrLeuLysValAlaLysProThrSerMetTyrMetAsnAsp 412
```

```
QY 598 GGCCTGATCCCC-----GTGAAAGCTCTCTGATGAGCCCTCACTTGGCAGTATGTG 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 GlyTyrAlaProProAspIleThrAsnSerSerGlnPro-----Val 427
QY 652 GATCAGTTGTGTCGGAATCTGGGGCGCTGCGAGCGCCTGGCTCCCTC----- 702
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 AspAsnHisValSerProSerSerPheLeuGlyGlnThrProAlaSerProAlaArgTyr 447
QY 703 -----GGAAATCGGGAA-----ACRAGGAGGAAGAGTCTTC 735
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 448 SerProValSerLysAlaValLeuGlyAspAspGluLeuThrArgGluProArgLysVal 467
QY 736 ATCAGCTGGTAGGCTCCCGAGGCTTGGCTGCAGCATTTCCAGCGCCCATCCAGAAG 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 ValLeuHisArgGlySerThrGlyLeuGlyPheAsnIleValGlyGlyGlu---AspGly 486
QY 796 CCGTGCATCTTTATCAGCCATGTG-----AAACCTGGCTCCTGTCTGCTGAGGTG 846
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 GluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu--- 505
QY 847 GGATGGAGATAGGGACCATGTCGAAGTCAATGCGCTCGACTTCTTAACCTGGAT 906
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 ---LeuArgLysGlyAspArgIleLeuSerValAsnSerValAspLeuArgAlaAlaSer 524
QY 907 CACAGGAGGCTGTAATATGCTGCTGAAATAGACCGCAGCTGACCATCTCCATTTAGCT 966
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 HisGluGlnAlaAlaAlaLeuLysAsnAla---GlyGlnAlaValThrIleValAla 543
QY 967 GCAGCTGCCGGGAGCTCTTTCATGACAGCCGGGAGCGGTGGCAGAGCGCGGAG--- 1023
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 544 GlnTyrArgProGluGluTyrSerArgPheGluAlaLysIleHisAspLeuArgGluGln 563
QY 1024 -----CGTGAGCTGCAGCGCGAGGAG 1044
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 564 MetMetAsnSerSerIleSerSerGlySerGlySerLeuArgThrSerGlnLys-ArgSe 583
QY 1045 CTTCTCATGCAGAGCGGTGGCGATGGAGTCCACAAAGATCCTCC----- 1090
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 583 rLeuTyrValArgAlaLeuPheAspTyrAspLysThrLysAspSerGlyLeuProSerG1 603
QY 1091 -----AGACAGCAGCAGGAGATG 1107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 603 nGlyLeuAsnPheLysPheGlyAspIleLeuHisValIleAsnAlaSerAspAspGluTr 623
QY 1108 G-----AGCGCAAGAGAGAA----- 1123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 623 pTrpGlnAlaArgGlnValThrProAspGlyGluSerAspGluValGlyValIleProSe 643
QY 1124 -AAGAAATTGCCAGAGCAGCAGAG-----AA 1152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 643 rLysArgArgValGluLysLysGluArgAlaArgLeuLysThrValLysPheAsnSerLy 663
QY 1153 AATGAGAGATACCGNAGAGAGATGGACAGATGTAGAGGAGGAAGAGTCTTAAGAG 1212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 663 sThrArgAspLysGlyGlnSerPheAsnAsp-----LysArgLysLysAsnLeuPheSe 681
QY 1213 C-----AATGGGAAGAGAGACTGGGGCTCAAGGAAGACTACTCTTG 1254
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 681 rArgLysPheProPheTyrLysAsnLysAspGlnSerGluGlnGluThrSerAspAla--- 700
QY 1255 CCTAAACCATCACTGCTGAGGTACACCCAGTACCCTTCGCAAGCCAAAGTATTCCG 1314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 701 -----AspGlnHisValThrSerAsnAlaSerAspSerG1 712
QY 1315 GAATATGAGGAGGCTTTGACCCCTACTCTATGTTCACCCAGCAGCAGATCATGGGAA 1374
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 712 uSerSerTyrArgGlyGlnGluGluTyr---ValLeuSerTyrGluProValAsnGlnG1 731
QY 1375 GGATGTCGGCTCCTACGATCAAGAAGGAGGATCTCTTAGACCTCGCCCTGGAAGCGG 1434
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 731 nGluValAsnTyrThrArg-----ProValIleIleLeuGlyPr 744
```

```
QY 1435 TGRGACTCCCCCATTTGGGAAGGTGCTGTTTCTGCTGTGTATGACCGG---GGACCTGC 1491
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 oMetLysAspArgIleAsnAspAspLeuIleSerGluPheProAspLysPheGlySerCy 764
QY 1492 TGAGCGCCATGTTGGCATTTCTGAAGGGAC---GAGATCATGGCAATCAACGCAAGAT 1548
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 764 sValProHisThrThrArgProLysArgAspTyrGluValAspGlyArgAspTyrHisPh 784
QY 1549 TGTGACAGACTACACCTGCTGAGCTGAGCTGCCTGCAGAAAGCCCTGGAAATCAGGG 1608
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 784 eValThrSerArgGluGlnMetGluLysAspIleGlnGluHisLys---PheIleGluAl 803
QY 1609 CGGAGACTTGATGCACCTTGTG-----GTTGCCGTCTGCCCCCCCAAGAGGATGACGA 1662
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 803 aGlyGlnTyrAsnAsnHisLeuTyrGlyThrSerValGlnSerValArgGluValAlaG1 823
QY 1663 TGAGCTGACCTTCTGCTGAAGTCCAAAAGGGAAACCAAAATTCACGCGTTAGGAAACAG 1722
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 823 yLysGlyLysHisCysIleLeuAspValSerGlyAsnAlaIleLysArgLeuGlnIleAl 843
QY 1723 TGAGCTCCGCCCCACCTCGTGAACACAAAGCCCTCGGACC 1762
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 843 aGlnLeuTyrProIleSerIlePheIleLysProLysSer 856
RESULT 6
A46431
tight junction-associated protein ZO-1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A46431
R:Itch, M.; Nagafuchi, A.; Yonemura, S.; Kitani-Yasuda, T.; Tsukita, S.;
J. Cell Biol. 121, 491-502, 1993
A:Title: The 220-kD protein colocalizing with cadherins in non-epithelial cells is id
py.
A:Reference number: A46431; MUID:93252986; PMID:8486731
A:Accession: A46431
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1745 <IT0>
A:Cross-references: GB:D14340; NID:g303709; PIDN:BAA03274.1; PID:d1003784; PID:g30371
A:Experimental source: F9 cells
A:Note: sequence extracted from NCBI backbone (NCBI:131200, NCBI:131201)
C:Superfamily: guanylate kinase homology; GLGF domain homology
F:27-106/Domain: GLGF domain homology <GLG1>
F:428-498/Domain: GLGF domain homology <GLG3>
F:645-794/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 1.33e-07 Length: 1745
Score: 236.00 Matches: 146
Percent Similarity: 37.63% Conservative: 102
Best Local Similarity: 22.15% Mismatches: 218
Query Match: 5.92% Indels: 193
DB: 2 Gaps: 32
US-09-502-945-5 (1-2162) x A46431 (1-1745)
QY 358 GTGCGCTGGACCGTCTGCACCGCAGAGCCCTCGCGTGTGTCGTGTCGTGGCTGGAG 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24 ValThrLeuHisArgAlaProGlyPheGlyPheGlyIleAlaIleSerGlyArgAsp 43
QY 418 -----TTTGGCTGTGGG-----CTCTTCATCTCCCACTCCTCAAGGCGGT 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 AsnProHisPheGlnSerGlyGluThrSerIleValIleSerAspValLeuLysGlyGly 63
QY 460 CAGCAGACACCGCTCGGCTCCAGGTAGGGACGAGATCGTCCGATCAATGATATATCC 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 ProAlaGluGly---GlnLeuGlnGluAsnAspArgValAlaMetValAsnGlyValSer 82
QY 520 ATCTCTCTCTGATCCCATGAGGAGGTCTATCAAC----- 552
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 MetAspAsnValGluHisAlaPheAlaValGlnGlnLeuArgLysSerGlyLysAsnAla 102
```

```

QY 553 -----CTCATTGCAACCAAGAACTGTGTCATCAAAAGTGAGACATCGGCTGATC 606
Db 103 LysIleThrIleArgArgLysLysValGlnIleProValSerHis----- 118
QY 607 CCGGTGAAAGCTCTCCGAT---GAGCCCTCACTTGGCAGTATGTGCATCAGTTGTG 663
Db 119 -----ProAspProGluProValSerAspAsnGluAspSerTyrAsp 133
QY 664 TCGGAATCTGGGGCGTGGAGGAGCCCTGGCTGCCCTGGAATCTGGGAAATCGGGAACAAGAG 723
Db 134 GluGluValHisAspProArgAlaGlyArgGlyAlaLeuAlaAsnArgArgSerGluLys 153
QY 723 ----- 723
Db 154 SerTrpAlaArgAspArgSerAlaSerArgGluArgSerLeuSerProArgSerAspArg 173
QY 724 -----AAGAAGTCTTCATCAGCCTGGTAGCTCCCGA 756
Db 174 ArgSerValAlaSerSerGlnProAlaLysProThrLysValThrLeuValLysSerArg 193
QY 757 -----GGCTTGGCTGCAGCATTTCCAGCGGCCCATCCAGCAAGCTGGC 801
Db 194 LysAsnGluGluTyrGlyLeu-----ArgProAla 203
QY 802 -----ATCTTTATCAGCCATGTGAACCTGGCTCCCTGTCTGCTGAGGTGGGA---TTG 852
Db 204 SerHisIlePheValLysLysLysSerLysSerLeuAlaAlaArgAspGlyAspIle 223
QY 853 GAGATAGGGGACCAAGTTGTCGAGTCAATGCGCTGCTCTTCTTAACCTGGATCACAAG 912
Db 224 GlnGluGlyAspValValLeuLysIleAsnGlyThrValThrGluAsnMetSerLeuThr 243
QY 913 GAGCTGTAATATGCTCAAAATAGCCGACCTGACCATCTCCATCTGAGTGCAGCT 972
Db 244 AsnAlaLysThrLeuIleGluArgSerLysGlyLysLeuLysMetValValGlnArgAsp 263
QY 973 GCGCGGAGCTGTTCATG----- 990
Db 264 GluArgAlaThrLeuLeuAsnValProAsnLeuSerAspSerIleHisSerAlaAsnAla 283
QY 991 ACAGCCGGGAGCGGTGGCAGAGCGCGGCGAG-----CGTGAG 1029
Db 284 SerGluArgAspAspIleSerGluIleGlnSerLeuAlaSerAspHisSerGlyArgSer 303
QY 1030 CTGCAGCGGAGGAGCTCTCATCCAGAGCGGCTG----- 1065
Db 304 HisAspArgProProArgArgSerGlnSerArgSerProAspGlnArgSerGluProSer 323
QY 1066 -----GCGATGGAGTCCAAACCAAGATCCTCCAGGAGCAGCAGAG 1104
Db 324 AspHisSerThrGlnSerProGlnGlnProSerAsnGlySerLeuArgSerArgGluGlu 343
QY 1105 ATGAGCGGCAAGAGGAAAGAAATTTCCCAAGAGCAGCAGAGGAAATAGAGATAC 1164
Db 344 GluArgMetSerLysProGlyAlaIleSerThrProValLysHisValAspAspHisPro 363
QY 1165 CGGAGGAGATGGACAGATT---GTAGCAGGAGGAAGAGATTAGACAGCAATGGGAA 1221
Db 364 ProLysAlaValGluGluValThrValGluLysAsnGluLysGlnThrProThrLeuPro 383
QY 1222 GAAGACTGGGGCTCAAAGGAACAGCTACTCTTGCTTAAACCATCATCTGCTAGGTACAC 1281
Db 384 Glu-----ProLysProValTyrAlaGlnValGly 393
QY 1282 -----CCAGTACCCCTTCGCAAGCCAAAGTGTTCGCGAAATATGA 1323
Db 394 GlnProAspValAspLeuProVal-----SerPro-SerAsp----- 405
QY 1324 GGAAGGCTTGACCCCTACTCTATGTTACCCAGACAGATCATGGGAGGATGTCG 1383
Db 406 ---GlyAlaLeuProAsnSerAlaHis---GluAspGlyIleLeuArgProSerMetLys 423
QY 1384 GCTCTACGATCAAGAGGAGGATCTTAGACCTGGCCCTGGAAGCGGTGTGGACTC 1443

```

```

Db 423 sLeuValLysPheArgLysGlyAspSerValGlyLeuArgLeuAlaGlyGlyAsnAsp-- 442
QY 1444 CCCATTGGGAAGTGGTCTGCTGCTGTATGACGGGAGCTGCTGAGCGGCATGG 1503
Db 443 ---ValGly---IlePheValAlaGlyValLeuGluAspSerProAlaAlaLysGlu-- 459
QY 1504 TGGCATTTGTAAGGGGACGAGATCATGGAATCAACGGCAAGATTGTACAGACTACAC 1563
Db 460 -GlyLeuGluGluGlyAspGlnIleLeuArgValAsnAsnValAspPheThrAsnIleI 479
QY 1564 CTTGGCTGAGGCTACGCTGCCCTGTCAGAGGCTGGAATCAGGCGGGGACTGGATCGA 1623
Db 479 eArgGluGluAlaValLeuPheLeu-----LeuAs 489
QY 1624 CTTGTGTTGCTGCTGCTGCCCCCAAGAGTATGACGATGAGCTGACCTTCTTGCTGAA 1683
Db 489 pLeu-----ProLys-----GlyGluGluValThrIleLeuAlaG 501
QY 1684 GTCCAAAGGGGAACCAAAATTCACGCGTTAGGAACAGCTGCCGCCCCCCTCGT 1743
Db 501 nLysLysLys-----AspValTyrArgArgIleVa 511
QY 1744 GAACACAAAGCCTCGGACGAGCTTGAG---AGAGGCCACATGACACACAGATGGCA 1800
Db 511 lGluSerAspValGlyAspSerPheTyrIleArgThrHisPheGluTyrGluLysGluSe 531
QY 1801 TCCTTGGGACCTGAATCTATCACCAGGAATCTCAAACTCCCTTTGGCCCTCAACAGGG 1860
Db 531 rProTyrGly-----LeuSerPheAsnLysG 540
QY 1861 CCAGATTAAGAAACAGCTCGGCGCCACTTTTGAAGGCCAATGTGAGGAAGGAGGAGCAGC 1920
Db 540 yGluValPheArgValValAspThrLeuTyrAsnGlyLysLeu---GlySerTrpLeuAl 559
QY 1921 CAGCGTTTCGGAGAGATCTCAAGGATCCAGACTCTCATCTCTTCTTCTTCTTCTTCT 1969
Db 559 aIleArgIleGlyLysAsnHisLysGluValGluArgGlyIleIleProAsnLysAsnAr 579
QY 1970 -----CTGCGCCAGTGAATTTGGTCTCTCCCA-----GCTTTGGGGGAC 2008
Db 579 gAlaGluGlnLeuAlaSerValGlnTyrThrLeuProLysThrAlaGlyGlyAsp 597

```

RESULT 7

T09599

postsynaptic density protein 95 - human

C:Species: Homo sapiens (man)

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000

C:Accession: T09599

R:Stathakis, D.G.; Hoover, K.H.; You, Z.; Bryant, P.J.

submitted to the EMBL Data Library, July 1998

A:Reference number: Z16761

A:Accession: T09599

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-767 <STA>

A:Cross-references: EMBL:083192; NID:g3318652; PID:g3318653

A:Experimental source: mammary

C:Genetics:

A:Gene: PSD95

C:Superfamily: discs-large tumor suppressor; GLGF domain homology: guanylate kinase h

F:208-286/Domain: GLGF domain homology <GLG>

F:478-536/Domain: SH3 homology <SH3>

F:578-755/Domain: guanylate kinase homology <GKI>

Alignment Scores:

Pred. No.:	1,24e-07	Length:	767
Score:	234.50	Matches:	144
Percent Similarity:	38.10%	Conservative:	104
Best Local Similarity:	22.12%	Mismatches:	210
Query Match:	5.88%	Indels:	193
DB:	2	Gaps:	34

A: Molecule type: mRNA
A: Residues: 1-724 <CHO>
A: Cross-references: GB:M96853; NID:g206454; PIDN:AAA4197L.1; PID:g206455
A: Experimental source: brain
C: Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase homology
F: 70-148/Domain: GLGF domain homology <GLG1>
F: 165-243/Domain: GLGF domain homology <GLG2>
F: 435-493/Domain: SH3 homology <SH3>
F: 535-712/Domain: guanylate kinase homology <GKI>

Alignment Scores:	
Pred. No.:	1,66e-07
Score:	232.50
Percent Similarity:	38.10%
Best Local Similarity:	22.12%
Query Match:	5.83%
DB:	2
	34
	724
	144
	104
	210
	193
	34

US-09-502-945-5 (1-2162) x JH0800 (1-724)

340	QY	TCCAGGAGCTGAAGGAGGTGGCTGGACCGTGTGCACCCCGAAGCCCTCGGCCTGAGT	399	Db	155	AspGluLysValMetGluIleLysLeuIleLys---	GlyProLysGlyLeuGlyPheSer	173	QY	400	GTGCGTGGTGGC-----CTGAGTGTGGCTGGCTGCTTCATCTCCAC	444	Db	174	IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys	193	QY	445	CTCATCAAGCGGTGACGAGCAGCGCTGGG---CTCCAGTAGGGACGAGATCGTC	501	Db	194	IleIleGluGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu	213	QY	502	CGGATCAATGGATATTCATCTCTCTCTGTACCCATGAGGAGTGCATCAACCTCATTCGA	561	Db	214	AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaAlaLeuLys	233	QY	562	---ACCAGAAACTGTGTCATCAAGTG-----AGACATCCGCCCTGATCCCC	609	Db	234	AsnThrTyrAspValValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp	253	QY	610	GTGAAAAGCTCTCCTGATGAGCCCCCTACACTGG---CAGTATGTGGATCAGTTTGTGTGC	666	Db	254	SerTyrAlaProAspIleThrThrSerTyrSerGlnHisLeuAspAsnGluIleSer	273	QY	667	GAATCTGGGGCGTGGAGCACCTGGGC-----SerrTyrLeuGlyThrAspTyrProThrAlaMetThrProThr	289	QY	697	-----TCCCGTGAATCCGGAA-----AACAAAG	720	Db	290	SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluLysPileProArg	309	QY	721	GAGAGAGAGCTCTTCATCAGCCCTGGTGTAGCTCCGAGCGCTTGGCTGCAGATTTCCACG	780	Db	310	GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly	329	QY	781	GGCCCCATCCAGAAGCCTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC	831	Db	330	GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp	348	QY	832	CTGCTCTGCTGAGTGGGATTTGAGATAGGGGACCATGTTGTCGAAGTCAATGGGCTGCAC	891	Db	349	LeuSerGlyGlyLysLeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp	366	QY	892	TTCTCTAACTGGATCACAAGAGGCTGTAATGTGCTCAAAATAGCCGCGCCTGACC	951	Db	367	LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr	385	QY	952	ATCTCCATTAGCTGCAGCTGCCGGGAGCTGTTTCATGCACAGACCGGGAGCGGCTGCCA	1011	Db	386	ValThrIleIleAlaGln-----TyrLysProGluGlyTyrSerArgPhe	400	QY	1012	GAGGCGCGCAGCGTGAAGTGCAGCGGACGAGGAGCTTCTCATGCAGAACCGCTCGCGATG	1071
-----	----	---	-----	----	-----	--------------------------------------	--------------------------	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	---	-----	----	-----	---	-----	----	-----	--	-----	----	-----	---	-----	----	-----	--	-----	----	-----	---	-----	----	-----	---	-----	----	-----	--	-----	----	-----	-----------------------------------	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	---	-----	----	-----	---	-----	----	-----	--	-----	----	-----	---	------	----	-----	--	-----	----	------	--	------

Db 588 heSerAlaIleValGluGlyAspSerPheGluGluIleTyrHisLysValLysArgValI 708
QY 1948 TCCAGACTCTCATT---CCTTCTCTGG 1973
Db 708 leGluAspLeuSerGlyProTyrIleTrp 717
RESULT 9
138757
homolog of Drosophila discs large protein, isoform 1 --human
C:Species: Homo sapiens (man)
C:Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
C:Accession: I38757
R:Lue, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.
Proc Natl Acad Sci U S A. 91, 9818-9822, 1994
A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila disc
A:Reference number: I38756; MUID:95024052; PMID:7937897
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-904 <RES>
A:Cross-references: EMBL:U13897; NID:g558437; PIDN:AAA50599.1; PID:g558438
C:Genetics:
A:Gene: GDB:DLG1
A:Cross-references: GDB:393278; OMIM:601014
A:Map position: 3q29-3q29
C:Superfamily: discs-large tumor suppressor; G1CF domain homology; guanylate kinase hom
C:Keywords: alternative splicing; duplication
F:229-307/Domain: G1CF domain homology <G1G1>
F:324-402/Domain: G1CF domain homology <G1G2>
F:588-646/Domain: SH3 homology <SH3>
F:715-892/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 1.94e-07 Length: 904
Score: 232.00 Matches: 82
Percent Similarity: 47.02% Conservative: 52
Best Local Similarity: 28.77% Mismatches: 103
Query Match: 5.82% Indels: 48
DB: 2 Gaps: 15
US-09-502-945-5 (1-2162) x I38757 (1-904)
QY 340 TCCAGGAAGCTGAAGAGGTGGCTGGACCGTCTGCACCCGAGGCGCTGGCGTGGT 399
Db 314 SerGluLysIleMetGluIleLysLeuLys---GlyProLysGlyLeuPheSer 332
QY 400 GTGCGTGGTGGC-----CTGGAGTTGGCTGGCGCTCTTCATCTCCAC 444
Db 333 IleAlaGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 352
QY 445 CTCATCAAGGCGGTGAGCAGCAGCGTGGG---CTCCAGTAGGGGAGCAGATCGTC 501
Db 353 IleIleGluGlyGlyAlaAlaHisLysAspGlyLysLeuGlnIleGlyAspLysLeuLeu 372
QY 502 CGGATCAATGGATATTCATCCCTCTGACCCATGAGGAGGTGATCAACCTCATTCGA 561
Db 373 AlaValAsnValCysLeuGluGluValThrHisGluGluAlaValThrAlaLys 392
QY 562 ---ACCAAGAAACTGTGTCCATCAAGTG-----AGACACATC 597
Db 393 AsnThrSerAspPheValTyrLeuLysValAlaLysProThrSerMetTyrMetAsnAsp 412
QY 598 GGCCTGATCCCC-----GTGAAGACTCTCTGTGATGAGCCCTCCTACTTGGCAGTATGG 651
Db 413 GlyTyrAlaProAspIleThrAsnSerSerSerGlnPro-----Val 427
QY 652 GATCAGTTGTGCGGAATCTGGGCGTGGCAGGCGAGCTGGCTCCCTC----- 702
Db 428 AspAsnHisValSerProSerPheLeuGlyGlnThrProAlaSerProAlaArgTyr 447
QY 703 -----GGAAATCGGGA---AACAGGAGGAAGAGTCTTTC 735
Db 448 SerProValSerLysAlaValLeuGlyAspGluIleThrArgGluProArgLysVal 467

QY 736 ATCAGCCTGGTAGGCTCCGAGCGCTTGGCTGCAGCATTTCAGCGGGCCCCCCTCCAGAAG 795
Db 468 ValLeuHisArgGlySerThrGlyLeuGlyPheAsnIleValGlyGlu---AspGly 486
QY 796 CTGCGCATCTTATCAGCCATGTG-----AAACCTGGCTCCCTGCTGCTGAGGTG 846
Db 487 GluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu--- 505
QY 847 GGATTGGAGTAGGCGACCATTTGCGAAGTCAGTGGCGTGCAGCTTCTTAACCTGGAT 906
Db 506 ---LeuArgLysGlyAspArgIleIleSerValAsnSerValAspLeuArgAlaAlaSer 524
QY 907 CACAGGAGGCTGTAATGTGCTGAAAAATAGCCGAGCGCTGACCATCTCCATGTAGCT 966
Db 525 HisGluGlnAlaAlaAlaAlaLeuLysAsnAla---GlyGlnAlaValThrIleValAla 543
QY 967 GCAGCTGGCGGAGGTGTTTCATGACAGCCGGAGCGGCTGCAGAGCGGCGGAGCGT 1026
Db 544 Gln-----TyrArgProGluGluTyrSerArgPheGluAlaLysIleHis 558
QY 1027 GAGCTGCAGCGGCGAGGAGCTTCTCAGAGCGGCTGGCGATGGATCCAAACAGATC 1086
Db 559 AspLeuArgGluGln-----MetMetAsnSerSerIleSerGlySerGlySerLeu 576
QY 1087 CTCAGGAGCAGCAG 1101
Db 577 ArgThrSerGlnLys 581
RESULT 10
S71625
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 13 - mouse
N:Alternate names: epidermal growth factor-binding protein; serine proteinase
C:Species: Mus musculus (house mouse)
C:Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 21-Jun-2002
C:Accession: S71625; S67987; I81210; I81209; S40290
R:Chida, D.; Kume, T.; Mukoyama, Y.; Tabata, S.; Nomura, N.; Thomas, M.L.; Watanabe,
FEBS Lett. 358, 233-239, 1995
A:Title: Characterization of a protein tyrosine phosphatase (RIP) expressed at a very
A:Reference number: S71625; MUID:95145716; PMID:7843407
A:Accession: S71625
A:Molecule type: mRNA
A:Residues: 1-2450 <CHI>
A:Cross-references: EMBL:D83966; NID:g1232103; PIDN:BAAL2158.1; PID:g1232104
A:Experimental source: strain DBA/2; cell line MEL 745A
R:Wolf, B.B.; Brown, M.D.
FEBS Lett. 376, 177-180, 1995
A:Title: Epidermal growth factor-binding protein activates soluble and receptor-bound
A:Reference number: S67987; MUID:96105375; PMID:7498536
A:Accession: S67987
A:Molecule type: protein
A:Residues: 1098-1102 <WOL>
A:Experimental source: submaxillary glands
R:Sato, T.; Irie, S.; Kitada, S.; Reed, J.C.
Science 268, 411-415, 1995
A:Title: FAP-1: a protein tyrosine phosphatase that associates with Fas.
A:Reference number: I59595; MUID:95232528; PMID:7536343
A:Accession: I81210
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1329-1354, 'K', 1356-1447, 'R', 1449-1454 <RES>
A:Cross-references: GB:I34582; NID:g806297; PIDN:AAC42056.1; PID:g806298
A:Accession: I81209
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1338-1354, 'K', 1356-1447, 'R', 1449-1454 <RE2>
A:Cross-references: GB:I34581; NID:g806295; PIDN:AAC42055.1; PID:g806296
R:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.
submitted to the EMBL Data Library, June 1993
A:Description: Assessment of the expression levels of murine protein-tyrosine phosph
A:Reference number: S40280
A:Accession: S40290
A:Molecule type: mRNA

A:Residues: 2256-2372 <HEN>
 A:Cross-references: EMBL:Z23059; NID:g438155; PIDN:CAAR0594.1; PID:g438156
 C:Genetics:

A:Gene: Ptpn13
 A:Map position: 5

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology;
 F:566-860/Domain: phosphoric monoester hydrolase; transmembrane protein; tyrosine
 F:1089-1165/Domain: GLGF domain homology <GLG1>
 F:1361-1437/Domain: GLGF domain homology <GLG2>
 F:1495-1574/Domain: GLGF domain homology <GLG3>
 F:1769-1840/Domain: GLGF domain homology <GLG4>
 F:1863-1937/Domain: GLGF domain homology <GLG5>
 F:2203-2422/Domain: protein-tyrosine-phosphatase homology <Ptp1>
 F:2374/Active site: Cys (phosphotyrosine intermediate) #status predicted
 F:2380/Binding site: substrate phosphate (Arg) #status predicted

Alignment Scores:
 Pred. No.: 6,02e-07 Length: 2450
 Score: 227.00 Matches: 146
 Percent Similarity: 32.23% Conservative: 106
 Best Local Similarity: 18.67% Mismatches: 192
 Query Match: 5.63% Indels: 338
 DB: 2 Gaps: 30

US-09-502-945-5 (1-2162) x S71625 (1-2450)

QY 308 AGGTGGAATATGATCAGCTGACCCCGCGCTCCAGGAAGCTGAGGCTGCTGG 367
 Db 1070 ArgtrpSerIleValSer--SerProGluArgGluIleThrLeuValAsnLeuLysA 1089
 QY 368 ACCGTCTCACCCGAGAGCTGCGCTGAGTGTGCGTGGCTGGAGTTGGC---- 423
 Db 1089 spProLysHis-----GlyLeuGlyPheGlnIleGlyGlyLysMetGlyArgL 1107
 QY 424 ----TGTGGCTCTTCATCTCCACCTCATCAAGGCGGTGAGGCTGCTGCTG 477
 Db 1107 euAspLeuGlyValPheIleSerAlaValThrProGlyGlyProAlaAspLeuAspGlyC 1127
 QY 478 --CTCCAGTGTGGGAGAGATCTCCGGATCATGATATTCATCTCCCTGACCC 535
 Db 1127 ysLeuLysProGlyAspArgLeuIleSerValAsnSerValSerLeuGluGlyValSerH 1147
 QY 536 ATGAGGAGGTCAACCTCATTCGA----- 561
 Db 1147 isHisAlaAlaValAspIleLeuGlnAsnAlaProGluAspValThrLeuValIleSerG 1167
 QY 561 ----- 561
 Db 1167 lnProLysGluLysProSerLysValProSerThrProValHisPheAlaAsnGlyMetL 1187
 QY 562 -----ACCAAGAAACTGTCTCCATCAA----- 585
 Db 1187 ysSerTyrThrLysLysProAlaTyrMetGlnAspSerAlaMetAspProSerGluAspG 1207
 QY 586 -----CTGACACACATCGCGCTGATCCCGTGAAAGCTCTCTG 625
 Db 1207 lnProTrpProArgGlyThrLeuArgHisIle-----ProG 1219
 QY 626 ATGAGCCCTCACTTGGCAGTATGTGGATCAGTTGTCTCGAATCTGGGGCGCTGCGA- 684
 Db 1219 luSerProPhe-----GlyLeuSerGlyGlyLeuArg 1230
 QY 685 --GGCAGCTGGGCTCCCTGGAAATCGG-----GAAACCAAGG 721
 Db 1230 luGlySerLeuSerSerGlnAspSerArgThrGluSerAlaSerLeuSerGlnSerGlnV 1250
 QY 722 AGAAGAGCTTTCATCAGCTGTAGCTCCCGAGCG----- 759
 Db 1250 alaAsnGlyPhePheAlaSerHisLeuGlyAspArgGlyTrpGlnGluProGlnHisSers 1270
 QY 759 ----- 759

Db 1270 erProSerProSerValThrThrLysValAsnGluLysThrPheSerAspSerAsnArgS 1290
 QY 759 ----- 759
 Db 1290 erLysAlaLysArgArgGlyIleSerAspLeuIleGluHisLeuAspCysAlaAspSerA 1310
 QY 759 ----- 759
 Db 1310 spLysAspAspSerThrTyrThrSerSerGlnAspHisGlnThrSerLysGlnGluProS 1330
 QY 759 ----- 759
 Db 1330 erSerSerLeuSerThrSerAsnLysThrSerPheProThrSerSerAlaSerProProL 1350
 QY 760 -----CTTGGCTGCAGCA 772
 Db 1350 ysProGlyAspThrPheGluValGluLeuAlaLysThrAspGlySerLeuGlyIleSerV 1370
 QY 773 TTTCCAGCGGC-----CCCATCCAGAACCTGGCATCTTTTATCAGCCATCTGAAC 823
 Db 1370 alThrGlyGlyValAsnThrSerValArgHisGlyGlyIleTyrValIleAlaIleleP 1390
 QY 824 CT---GGCTCCCTCTCTGCTGAGTGGATGGAGATAGGGACAGAGTGTCCGAAGTCA 880
 Db 1390 roLysGlyAlaAlaGluSerAspGlyArgIleHisLysGlyAspArgValLeuAlaValA 1410
 QY 881 ATGGCGTCGACTTCTTAACCTGGATCAGAGGAGCTGTAATGTCTGCTGAAATAATGCC 940
 Db 1410 snGlyValSerLeuGluGlyAlaThrHisLysGlnAlaValGluThrLeuArgAsnThrG 1430
 QY 941 GCAGCTGAGCATCTCCATTTAGTTCAGCTGGCGGGAGCTGTTCTCATCAGACAGCCGG 1000
 Db 1430 lyGlnVal---ValHisLeuLeuLeuLysGlyGln-----ValProThrSerArg 1447
 QY 1001 AGCGGTGGCAGAGCGCGGCGAGCTGAGCTGCAGCGGAGGAGCTTCTCATGCAGAGC 1060
 Db 1447 luGlnAspProAlaGlyProGlnSerPro----- 1456
 QY 1061 GGCTGGCGATGGAGTCCCAACAAGATCTCCAGGAGCAGGAGATGAGCGGCAAGGA 1120
 Db 1457 -----ProProAspGlnAspAlaGlnArgGlnAlaP 1467
 QY 1121 GAAAGAAATTTGCCAGAGGCGAGAGAGAAATGAGAGATACCGGAGAGATGGAAC 1180
 Db 1467 roGluLysValAlaLysGlnThrPro-----HisValLysAspTyrSerP 1482
 QY 1181 AGATTGTAGAGGAG-----GAAGAGAAGTTTAAAGAAGCAATGGGAAGAGAGCTGG 1231
 Db 1482 heValThrGluAspAsnThrPheGluValLysLeuPheLys-AsnSerSerGlyLeuGly 1501
 QY 1232 GCTCAAGGAGACAGCTACTCTGCTTAAACCATCATCTGCTGAGGTACACCCAGTACCC 1291
 Db 1502 -----PheSer 1503
 QY 1292 TTCGCAAGCCAAAGTGTATTTCCGGAATATGAGGAAGCTTTGACCCCTACTCTATGTT 1351
 Db 1504 PheSerArgGluAspAsn----- 1510
 QY 1352 ACCCCAGAGCAGATCATCGGGAAGGATGTCGCGCTCTACGCAATCAAGAGAGGAGATCC 1411
 Db 1511 IleProGluGlnIleAsnGlySer-----IleValArgValLysLys----- 1524
 QY 1412 TTAGACCTGGCCCTGGGAAGCGGTGTGGACTCCCCCATTTGGGAAGGTGGTCTGTTCTGCT 1471
 Db 1524 ----- 1524
 QY 1472 GTGTATGAGCGGGAGCTGCTGAGCGCATGTGTGGCATTGTGAAGGGGACGAGATCATG 1531
 Db 1525 LeuPheProGlyGlnProAlaAlaGluSerGlyLysIleAspValGlyAspValIleLeu 1544
 QY 1532 GCAATCAACGGCAAGATTGTGCAGACACTACACCTGGCTGAGGCTGACGCTGCCCTGCAG 1591
 Db 1545 LysValAsnGlyAlaProLeuLysGlyLeuSerGlnGlnAspValIleSerAlaLeuArg 1564


```

Db 682 lThrSerAsnAlaSerAspSerGluSerSerTyrHisGluTyrGlyCysSerLysGlyG1 702
QY 1267 ACTGCTGAGGTACAC---CAGTACCCTTCGCAAGCAA----- 1303
Db 702 yGlnGluGluTyrValLeuSerTyrGluProValAsnGlnGluValAsnTyrThrAr 722
QY 1304 ----- 1311
Db 722 gProValIleLeuGlyProMetLysAspArgValAsnAspAspLeuIleSerGluPh 742
QY 1312 C----- 1326
Db 742 eProAspLysPheGlySerCysValProHisThrThrArgProLysArgAspTyrGluVa 762
QY 1327 AGGCTTTGACCCCTACTCTATGTTCCACCCAGACAGATCATGGGAAAGGATGTCGGCT 1386
Db 762 lAspGlyArgAspTyrHisPheValThrSerArgGluGlnMetGluLysAspIleGlnG1 782
QY 1387 CCTACGCATCAAGAAGGAGGATCCTTAGACCTGGCCCTGGGAAGC---GGTGTGGACTC 1443
Db 782 uHisLysPheIleGluAlaGlyGlnTyrAsnAsnHisLeuTyrGlyThrSerValGlnSe 802
QY 1444 CCCCATGGGAAGGTGCTCTCTCTGCTGTATGAGCGGGAGCTGCTGACGGCATGG 1503
Db 802 r-----ValargAlaValAlaGluLysGly-----LysHisCy 813
QY 1504 TGGCATTTGTAAGGGACGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACAC 1563
Db 813 s-----IleLeuAspValSerGlyAsnAlaIleLysArgLeuG1 826
QY 1564 CTGGGCTGAG 1573
Db 826 nIleAlaGln 829

RESULT 12
T15617
hypothetical protein C25F6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15617
R:Bentley, D.
submitted to The EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid C25F6.
A:Reference number: Z18377
A:Accession: T15617
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1131 <BEN>
A:Cross-references: EMBL:U39742; NID:g1049455; PID:g1049459; PIDN:AAA80434.1; CESP:C25F6
C:Genetics:
A:Gene: CESP:C25F6.2
A:Introns: 105/1; 210/3; 283/3; 316/1; 346/2; 463/3; 566/1; 722/2; 897/1; 991/2; 1032/1;

Alignment Scores:
Pred. No.: 5,3e-07 Length: 1131
Score: 226.00 Matches: 130
Percent Similarity: 37.44% Conservative: 92
Best Local Similarity: 21.92% Mismatches: 200
Query Match: 5.67% Indels: 171
DB: 2 Gaps: 26

US-09-502-945-5 (1-2162) x T15617 (1-1131)
QY 151 AATGATGACAGAGGACTATCTCTATGATGTGCTGCGAATGTACCACGACCATGGAC 210
Db 243 AsnSerSerSerHisThrLeuHisGluArgGlnArg-----GlnThrSerHis 259
QY 211 GTGGCCGTGCTCGTGGGAGACCTGAAGCTGGTTCATCAATGAACCCAGCCGCTGCTCCTG 270
Db 260 AspGlyThrTyrArgGluThrThrArgThrValAspThrProSerGlyLeuGlu--- 278
QY 271 TTGTATGCCATTGCGCCCGCTGATCCCACTGAAGCACCAGGCTGAATATGATGATGATGACC 330

```

```

Db 279 -----ArgArgValValGluHisThrGlyValIle 288
QY 331 CCCCGCGCTCCAGGAAG---CTGAAGAGGTGCGTCTGACCGCTGTCACCCCGAA 384
Db 289 AspAspHisGlyArgLysTyrGluLeuGluAsnIleValLeuGluLysGlyHis---Thr 307
QY 385 GCCTCGGCTCAGTGTGCTGCTGCTG---GAGTTTGGCTGTGGG 429
Db 308 GlyLeuGlyPheSerIleThrGlyMetAspGlnProThrGluAspLysAspThrSer 327
QY 430 CTCTTCATCTCCACCTCATCAAGCGGTGAGCAGACAGCGTCCGG---CTCCAGGTA 486
Db 328 IleTyrValThrAsnIleGluGlyAlaAlaLeuAlaAspGlyArgMetArgLys 347
QY 487 GGGGACGAGATCGTCGGATCAATGATATTCATCTCTCTGTACCATGAGGAGGTC 546
Db 348 AsnAspIleIleThrAlaValAsnAsnThrAsnCysGluAsnValLysHisGluValAla 367
QY 547 ATCAACCTCATTCGAACCAAGAAACTGTG---TCCATCAAGTGTAGA----- 591
Db 368 ValAsnAlaLeuLysSerSerGlyAsnValValSerLeuSerLeuLysArgArgLysAsp 387
QY 591 ----- 591
Db 388 GluAlaPheLeuProIleGlyAsnPheGlyGlySerThrSerTyrLeuArgSerGly 407
QY 592 -----CACATCGGCTGATCCCC 609
Db 408 ValThrProSerValSerAlaGlyAsnLeuGlnHisAlaIleHisSerProSerAlaPro 427
QY 610 GTGAAGAGCTCTCTGATGAGCCCTCACTTGGCAGTATGTGGATCAGTTTGTCTCGGAA 669
Db 428 IleHisProProProProValHisGlySerLeuSerGlnLeu----- 444
QY 670 TCTGGGCGCTCGAGGAGCTGGGCTCCCTGGAAATCGGAAACAGGAGGAAG 729
Db 445 -----SerValGlyGln-----TyrArgSerThrArgProAsnThr 456
QY 730 GTCCTTCATCAGCTGGTA---GGTCCCGAGCGCTGCTGCGACGATTTCCAGCGCCCC 786
Db 457 SerValIleAspLeuValGlyAlaArgGlyLeuGlyPheSerIleAlaGlyGln 476
QY 787 ATCCAGAAG-----CCTGGCATCTTTATCAGCCAT---GTCAAACTCTGG 828
Db 477 GlyAsnGluHisValLysGlyAspThrAspIleTyrValThrLysIleIleGluGly 496
QY 829 TCCCTGTCTGTGAGTGGATGGAGTAGGGACAGATTGCGAAGTCAATGGCGCT 888
Db 497 AlaAlaGluLeuAspGlyArgLeuArgValGlyAspLysIleLeuGluValAspHisHis 516
QY 889 GACTTCTCACTGATCAACAAGGAGCTGTAATGCTGTAATAATAGCCGAGCGCTG 948
Db 517 SerLeuIleAsnThrThrHisGluAsnAlaValAsnValLeuLysAsnThrGlyAsnArg 536
QY 949 ACCATCTCCATTGTAGTGTGCGGAGCTGCTCATGACAGACCGGCGGCGCTG 1008
Db 537 Val-----ArgLeuLeuIle-GlnGlnGly----- 544
QY 1009 GCAGAGCGCGGACGCTGAGCTGCGCGGAGGAGCTTCTCATGACAGAGCGCTGCG 1068
Db 545 -----ThrGlyAlaIlePheAsnAspSerAlaSerGlnG1 556
QY 1069 ATGAGTCCACACAGATCTCCAGGACGACGAGGATGGAGCGGCAAGGAGAGAA 1128
Db 556 nPheMetProThr----- 560
QY 1129 ATTCCCGACAGAGCGACGAGGAAATGACAGATACCGGAGGAGATGGAAACAGATTGTA 1188
Db 561 -ThrProIleLeuArgProSerSerValGlnAsp-----TyrAsnArgSerG1 576
QY 1189 GAGGAGGAAGAGATTTAAGAACCAATGGAGAACACTGGGCTCAAAGGACACGTA 1248
Db 576 nMetGlySerGlnSerHisLeuSerTyrGly-----GlyProLeuAsnThrSerty 593

```

```
QY 1249 CTCTTGCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTGA 1308
    |||||
Db 593 rSer-----SerGlnAlaProIleAlaIleProLeuG1 604
QY 1309 TTTCCTCGGAATATGAGGAAGGCTTTTGACCCCTACTCTATGTTCCACCCAGCAGATCAT 1368
    |||||
Db 604 uProArg-----ProValGlnLeu-- 610
QY 1369 GGGGAAGGATGTCGGGCTCCTACGCATCAAGAAGAGGAGGATCCTTAGACCTGGCCCTCGGA 1428
    |||||
Db 611 -----ValLysGlyGlnAsnGlyLeuGlyPheAsnIleVa 622
QY 1429 AGCGGGTGTGGATCCCTCCCATTTGGGAAGGTGGCTGTTCTGCTGCTGTATGAGCGGGAGC 1488
    |||||
Db 622 lGlyGlyGluAspAsnGlu-----ProIleTyrIleSerPheValLeuProGlyGlyVa 640
QY 1489 TGTGAGCGGCATGTCGTCATGTGAAGGGGAGGAGATCATGGCAATCAACGGCAAGAT 1548
    |||||
Db 640 lAlaAspLeuSerGlyAsnValLysThrGlyAspValLeuGluValAsnGlyValVa 660
QY 1549 TGTGACAGACTACACCTGCGTGGCTGAGCTGCGCTGCCCTGCAGAAAGCCCTGGAATCAGGG 1608
    |||||
Db 660 lLeuArgAsnAlaThrHisLysGluAlaAlaGluAlaLeuArgAsnAla----- 676
QY 1609 CGGGAGCTGATCCACCTTGTGTTGCTGCGCTGCCCTCCCAAGAGGATGATGACGATGAGCT 1668
    |||||
Db 677 -GlyAsnProValTyrLeuThrLeuGlnTyr---ArgProGlnGluTyrGln----- 692
QY 1669 GACCTCTTCTGTAAGTCCAAAGGGGAAACCAATACACCGCTTAGGAAACAGTGAAGCT 1728
    |||||
Db 693 -----IlePheGluSerLys-----IleGluLysLeuArgAsnAspValI1 706
QY 1729 CCGGCCCCACCTGTAACACAAACCCCTGCAGCCAGC 1765
    |||||
Db 706 eAlaGlnSerArgMetGlyThrLeuSerArgLysSer 718

RESULT 13
A47747
tight junction protein ZO-1 - human
C:Species: Homo sapiens (man)
C:Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
C:Accession: A47747
R:Willott, E.; Balda, M.S.; Fanning, A.S.; Jameson, B.; Van Itallie, C.; Anderson, J.M.
Proc. Natl. Acad. Sci. U.S.A. 90, 7834-7838, 1993
A:Title: The tight junction protein ZO-1 is homologous to the Drosophila discs-large tum
A:Reference number: A47747; MUID:93361541; PMID:8395056
A:Accession: A47747
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1736 <WIL>
A:Cross-references: GB:L14837; NID:g292937; PIDN:AAA02891.1; PID:g292938
C:Superfamily: guanylate kinase homology; GLGF domain homology
C:Keywords: alternative splicing; membrane protein; phosphoprotein
F:15-94/Domain: GLGF domain homology <GLG1>
F:181-248/Domain: GLGF domain homology <GLG2>
F:416-486/Domain: GLGF domain homology <GLG3>
F:633-782/Domain: guanylate kinase homology <GKI>

Alignment Scores:
Pred. No.: 7,22e-07 Length: 1736
Score: 225.00 Matches: 138
Percent Similarity: 37.88% Conservative: 109
Best Local Similarity: 21.17% Mismatches: 227
Query Match: 5.64% Indels: 179
DB: 2 Gaps: 28

US-09-502-945-5 (1-2162) x A47747 (1-1736)

QY 358 GTGCGTCTGGACCGTGTGCACCCCGAAGCCCTCGGCGCTGAGTGTGGGTGGTGGCTGGAG 417
    |||||
Db 12 ValThrLeuHisArgAlaProGlyPheGlyPheGlyIleAlaIleSerGlyGlyArgAsp 31
```

```
QY 418 -----TTTGGCTGTGG-----CTCTTCATCTCCCACTCATCAAAAGCGGT 459
    |||||
Db 32 AsnProHisPheGlnSerGlyGluThrSerIleValIleSerAspValLeuLysGlyGly 51
QY 460 CAGGCAGACAGCGTCCGGCTCCAGGTAGGGAGAGATCTCCGGATCAATGGATATTC 519
    |||||
Db 52 ProAlaGluGly---GlnLeuGlnGluAsnAspArgValAlaMetValAsnGlyValSer 70
QY 520 ATCTCTCTCTGTACCATGATGAGGAGTCAACACCTCATTCGA---ACCAAGAAACATGTG 576
    |||||
Db 71 MetAspAsnValGluHisAlaPheAlaValGlnGlnLeuArgLysSerGlyLysAsnAla 90
QY 577 TCCATCAAAAGTGAACACATCGGC-----CTCATCCCGTGAAGAGCTCTCTCTGTATGAG 630
    |||||
Db 91 LysIleThrIleArgArgLysLysValGlnIleProValSerArgProAspProGlu 110
QY 631 CCCTCCTCCTGCGATGATGTCAGTTCGTGTCGGAATCTCGGGGGCTGCGAGGAGC 690
    |||||
Db 111 ProValSerAspAsnGluAspSerTyrAspGluGluIleHisAspProArgSerGly 130
QY 691 CTGGGCTCCCTCGAATCGGAAACACAGGAG----- 723
    |||||
Db 131 ArgSerGlyValValAsnArgArgSerGluLysIleTrpProArgAspArgSerAlaSer 150
QY 723 ----- 723
    |||||
Db 151 ArgGluArgSerLeuSerProArgSerAspArgSerValAlaSerSerGlnProAla 170
QY 724 AAGAAGTCTTCATCAGCTGGTGGTCCCGA-----GGCCTTGGCTCGACAT 774
    |||||
Db 171 LysProThrLysValThrLeuValLysSerArgLysAsnGluGluTyrGlyLeuArgLeu 190
QY 775 TCAGCGGCCCCATCCAGAAAGCTGTCATCTTATCAGCCATCTGAAACCTGCTCCCTG 834
    |||||
Db 191 AlaSer-----HisIlePheValLysGluIleSerGlnAspSerLeu 204
QY 835 TCTGCTCAGGTGGGA---TTGGAGATAGGGGACACAGATTCTCGAAGTCAATGGCTCGAC 891
    |||||
Db 205 AlaAlaArgAspGlyAsnIleGlnGluGlyAspValValLeuLysIleAsnGlyThrVal 224
QY 892 TTCTCTAACTTGGATCACAGGAGGCTGTAATGTGCTGAAATATAGCCCGCCTGACC 951
    |||||
Db 225 ThrGluAsnMetSerLeuThrAspAlaLysThrIleGluArgSerLysLysLeu 244
QY 952 ATCTCCATTTAGTGTGAGCTGGCGGGAGCTGTTC-----ATGACAGAC 996
    |||||
Db 245 LysMetValValGlnArgAspGluArgAlaThrLeuLeuAsnValProAspLeuSerAsp 264
QY 997 CGGAGCGGCTGCGAGAGCGGCGCAGCTGTCAGCTGCAGCGCAGGAGCTTCTCATGCAG 1056
    |||||
Db 265 SerIleHisSerAlaAsnAlaSerGluArgAsp----- 275
QY 1057 AAGCGCTGGGATGGAGTCCACAAAGATCTCCAGGAGCAGGAGATGGAGCGGCAA 1116
    |||||
Db 276 -----AspIleSerGluIleGlnSerLeuAlaSerAsp-HisSerGlyLar 290
QY 1117 AGGAAAGAAATGTCAGAGCAGCAGAGGAGGAAATAGAGATACCGGAGGAGATG 1176
    |||||
Db 290 gSerHisAspArgProArgArgSerArgSerArgSerProAsp----- 305
QY 1177 GAACAGATTTAGAGGAGGAGAGAGAGTTTAAAGCAATGGGGAAGAGACTGGGGCTCA 1236
    |||||
Db 306 -----GlnArgSerGluProSerAspHisSerArgHisSerProG1 319
QY 1237 AAGGAACAGCTACTCTTGCCTA-----AAACCATCACT 1269
    |||||
Db 319 nGlnProSerAsnGlySerLeuArgSerArgAspGluArgIleSerLysProGlyAl 339
QY 1270 GCTGAGGTACACC-----AGTACCCCTTCGCAAGCCCAAGTAT 1311
    |||||
Db 339 aVal-SerThrProValLysHisAlaAspHisThrProLysThrValGluGluValT 359
QY 1312 CCGGAAATATGAGGAAGGCTTTTGACCCCTACTCTATGTTTCCACCCAG----- 1358
```

```
Db 359 hrValGluArgAsnGluLysGlnThrProSerLeuProGluProLysProValTyrAlaG 379
QY 1359 -----AGCAGATCATGGG----- 1372
Db 379 lnValGlyAsnGlnMetTrpIleTyrLeuSerValHisLeuMetValSerTyrLeuIleG 399
QY 1373 -----AAGATGTCGGGCTCTACGCATCAAGAAGAGG 1406
Db 399 lnLeuMetLysMetGlyPheLeuArgProSerMetLysLeuValLysPheArgLysGlyA 419
QY 1407 GATCCTTAGACCTGGCCCTGGAAGCGGTGGTGGACTCCCGCATTTGGGAGGTGGTCGTTT 1466
Db 419 sPSerValGlyLeuArgLeuAlaGlyLysAsp-----ValGly-----IlePheValA 436
QY 1467 CTGCTGTGTATGAGGGGAGCTGTGAGCGCATGTGTGAAGGGAGCGAGGAGA 1526
Db 436 laglyValLeuGluAspSerProAlaAlaLysGlu--GlyLeuGluGluGlyAspGlnI 455
QY 1527 TCATGCCAATCAACGGCAGATTGTGACAGACTACACCTGCTGAGCTGACGCTGCC 1586
Db 455 leLeuArgValAsnAsnValAspPheThrAsnIleIleArgGluGluAlaValLeuPheL 475
QY 1587 TCAGAAAGCCCTGGAATCAGGCGGGAGCTGGATCGACCTTGTGTGCTGCCCTGCC 1646
Db 475 eu-----LeuAspLeu-----p 479
QY 1647 CAAGGAGTATGACGATGAGCTGACCTTCTTGTGAAGTCCAAAGGGGAAACCAAAATTC 1706
Db 479 roLys-----GlyGluGluValThrIleLeuAlaGlnLysLys----- 492
QY 1707 AGCCCTTAGGAACAGTGAAGCTCGGCGCCCGCCCTCGTGAACACAAAGCTCGGACCGCC 1766
Db 493 -----AspValTyrArgIleValGluSerAspValGlyAspSerp 507
QY 1767 TTGAG---AGAGGCCACATGACACACACAGATGCATCTTGGACCTGATCATCATCAC 1823
Db 507 heTyrIleArgThrHisPheGluTyrGluLysGluSerProTyrGly----- 522
QY 1824 CCAGGAATCTCAAACTCCCTTTGGCCCTGAACCGGCGCAGATAAGAACAGCTCGGCGCC 1883
Db 523 -----LeuSerPheAsnLysGlyGluValPheArgAlaValAspT 536
QY 1884 ACTTTTTCAGAGCAATGTGGAGAAAGGAGGAGCAGCCGCTTTGGGAGAAATCTCA 1943
Db 536 hrLeuTyrAsnGlyLysLeu---GlySerTrpLeuAlaIleArgIleGlyLysAsnHisL 555
QY 1944 AGGATCCAGACTCTCATTTCTCTCT-----CTGGCCCGAGTGA 1982
Db 555 ysGluValGluArgGlyIleIleProAsnLysAsnArgAlaGluGlnLeuAlaSerValG 575
```

RESULT 14

Tl0811
channel associated protein of synapse 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: Tl0811
R:Irle, M.; Hata, Y.; Takai, Y.
submitted to the EMBL Data Library, April 1996
A:Description: Cloning of new isoforms of PSD-95/SAP90 related genes.
A:Reference number: Z17166
A:Accession: Tl0811
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-852 <IRI>
A:Cross-references: EMBL:U53368; NID:g1517939; PID:g1517940
C:Genetics:
A:Note: PSD-95/SAP90-related gene 2 (chapsyn isoform2)
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase homol

F:198-276/Domain: GLGF domain homology <GLG>
F:543-601/Domain: SH3 homology <SH3>
F:663-840/Domain: guanylate kinase homology <GKI>

Alignment Scores:
Pred. No.: 1,51e-06 Length: 852
Score: 218.50 Matches: 120
Percent Similarity: 39.50% Conservative: 85
Best Local Similarity: 23.12% Mismatches: 188
Query Match: 5.48% Indels: 126
Gaps: 24

US-09-502-945-5 (1-2162) x Tl0811 (1-852)

```
QY 307 CAGGTGGAATATGATCAGCTGACCCCGCGCTCCAGGAAGCTGAAGGAGTGGCTGTG 366
Db 91 GluIleGluTyrGlu-----PheGluGluIleThrLeu 101
QY 367 GACCGTCTGCACCCCGAAGGCGCTCGGCTGAGTGTGCTGGCTGGCTGGAG----- 417
Db 102 GluArgGlyAsn---SerGlyLeuGlyPheSerIleAlaGlyGlyThrAspAsnProHis 120
QY 418 -----TTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGCGGTTCAGGCAGACAGC 471
Db 121 IleGlyAspAspProGlyIlePheIleThrLysIleIleProGlyGlyAlaAlaAlaGlu 140
QY 472 GTCGGG---CTCCAGGTAGGAGCAGATCTCGGATCAATGATTCATCTCTCTCC 528
Db 141 AspGlyArgLeuArgValAsnAspCysIleLeuArgValAsnGluValAspValSerGlu 160
QY 529 TGTACCCATGAGGAGTCACTCAACCTCATTCGTA---ACCAAGAAAACTGTCTCCATCAA 585
Db 161 ValSerHisSerLysAlaValGluAlaLeuLysGluAlaGlySerIleValArgLeuTyr 180
QY 586 GTGAGACACATCGGCGCTGATCCCGTGAAAGCTCTCTGTATGAGCCCTCCTTGGCAG 645
Db 181 ValArgArg----- 183
QY 646 TATGTGGATCAGTTGTGTGCGAATCTGGGGCGCTGCGAGCGACCTTGGGCTCCCTTGA 705
Db 183 ----- 183
QY 706 AATCGGAAACAAAGAGAAGAGTCTTTCATCAGCTGGTA---GGCTCCCGAGGCGCTT 762
Db 184 ArgArgProIleLeuGluThrValGluIleLysLeuPheLysGlyProLysGlyLeu 203
QY 763 GGCTGACGATTTCCAGCGGCGCCATCCAGAG-----CCTGGC-----ATCTTT 807
Db 204 GlyPheSerIleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyr 223
QY 808 ATCAGCCATGTGAACCTGGCTCCCTGCTGCTGAGTGGGA---TTGGAGATAGGGGAC 864
Db 224 ValThrLysIleIleAspGlyGlyAlaAlaGlnLysAspGlyArgLeuGlnValGlyAsp 243
QY 865 CAGATTCTCGAAGTCAATCGCTGCGACTTCTTAACCTGGATCACAAGGAGGTGTAAT 924
Db 244 ArgLeuLeuMetValAsnAsnTyrSerLeuGluGluValThrHisGluAlaValAla 263
QY 925 GTGCTGAAAATAGCCCGACGCTGACCATCTCCATTCCTAGCTGCGAGTGGCGGAGCTG 984
Db 264 IleLeuLysAsnThr---SerAspValValTyrLeuLysValGlyLysProThrThrIle 282
QY 985 TTATGACACACCGGAGCGGCTGGCAGAGCGGCGGCGGCGAGCTGAGCTGACGCGCAGGAG 1044
Db 283 TyrMetThrAspProTyrGlyProProAspIle-----ThrHisSerTyrSer 298
QY 1045 CTCTCATGCAGAAAGCGCTGGCGATGGAGTCCCAACAAGATCTCCAGGAGCAGCAGGAG 1104
Db 299 ProProMetGluAsnHisLeuLeuSerGlyAsnGlnGlyThrLeuGluTyrLysThrSer 318
QY 1105 ATG-----GAGCGGCAAGAGGAAAAAAGAAATTCGCCAGAGGAGCAGCAGGAGAAAT 1155
Db 319 LeuProProIleSerProGlyArgTyrSerProIleProLysHisMetLeuValGluAsp 338
```



```

Db 364 rSerProValGluCysAspLysSerPheLeuLeuSerAlaProTyrSerHisTyrHisLe 384
QY 1327 AGGCTTTGACCCCTACTATGTTACCCAGAG----- 1360
    ||| ||| |||
Db 384 uGlyLeuLeuProAspSerGluMetThrSerHisSerGlnHisSerThrAlaThrArgGl 404
QY 1361 -----CAGATCATGGGAAGGATGTCGGCTCTCAGG 1392
    : : : ||| : : :
Db 404 nProSerMetThrLeuGlnArgAlaValSerLeuGluGlyGluProArgLysValValLe 424
QY 1393 CATCAAGAGAGGATCCCTAGACCTGGCCCTGGAGCGGTGTGGACTCCCCCATGG 1452
    ||| ||| : : : ||| ||| |||
Db 424 uHisLysGlySerThrGlyLeuGlyPheAsnIleValGlyGlyGluAspGlyGlu----- 442
QY 1453 GAAGGTGGTCGTTCTGCTGTATGAGGGGAGCTGCTGAGCGGCATGGTGGCATGCT 1512
    : : : ||| ||| : : : ||| ||| : : :
Db 443 -GlyIlePheValSerPheIleuAlaGlyGlyProAlaAspLeuSerGlyGluLeuGl 462
QY 1513 GAAAGGGACGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCCCTGGCTGA 1572
    : : ||| ||| : : : ||| ||| : : :
Db 462 nArgGlyAspGlnIleLeuSerValAsnGlyIleAspLeuArgGlyAlaSerHisGluGl 482
QY 1573 GGCTGACGCTGCCCTGCAGAGCCCTGGNAATCAGCGGGGAGCTGGATCGACCTTGTGCT 1632
    : ||| ||| ||| ||| : : |||
Db 482 nAlaAlaAlaAlaLeu-----LysGlyAlaGlyGlnThrValThrIleIl 497
QY 1633 TGCCGTCTGCCCCCAAGAGGAGTATGACGATGAGCTGACCTTCTTGTGCTGAAGTCCAAAG 1692
    : ||| ||| : : : |||
Db 497 eAlaGlnTyrGlnProGluAspTyr-----AlaArgph 508
QY 1693 GGGAAACCAAAATTCACGCTTAGGAAACAGTGAGTCGGGCCCCACCTCGTGAAC 1747
    : : ||| ||| : : ||| |||
Db 508 eGluAlaLysIleHis-----AspLeuArgGluGlnMetMetAsn 521

```

Search completed: March 21, 2003, 13:06:25
Job time : 72.5119 secs

Result	Query	Score	Match	Length	DB	ID	Description
1	2648.5	66.4	552	1	A175_HUMAN	Q9Y6n9	homo sapien
2	272	6.8	905	1	ZO3_MOUSE	O9dxy1	mus musculus
3	271	6.8	933	1	ZO3_HUMAN	O95049	homo sapien
4	255.5	6.4	898	1	ZO3_CANFA	O62683	canis famill
5	249.5	6.3	849	1	DIG3_MOUSE	P07175	mus musculus
6	243.5	6.3	849	1	DIG3_RAT	P62936	rattus norv
7	236	5.9	1745	1	ZO1_MOUSE	P39447	mus musculus
8	234.5	5.9	767	1	DIG4_HUMAN	P78352	homo sapien
9	232.5	5.8	724	1	DIG4_RAT	P31016	rattus norv
10	232.5	5.8	817	1	DIG3_HUMAN	Q92796	homo sapien
11	232	5.8	904	1	DIG1_HUMAN	Q12959	homo sapien
12	229.5	5.8	724	1	DIG4_MOUSE	Q62108	mus musculus
13	226	5.7	911	1	DIG1_RAT	Q62696	rattus norv
14	225	5.6	1736	1	ZO1_HUMAN	Q07157	homo sapien
15	215.5	5.4	852	1	DIG2_RAT	Q63622	rattus norv
16	213.5	5.4	870	1	DIG2_HUMAN	Q15700	homo sapien
17	204.5	5.1	2485	1	PTND_HUMAN	Q12923	homo sapien
18	195.5	4.9	2161	1	SHK1_HUMAN	Q9Y566	homo sapien

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announcement/)
CC or send an email to license@sib-sib.ch).

CC
DB EMBL: AF039700: AAC18049.1: ALT FRAME.

Db	161	SerIleLysValArgHisIleGlyLeuIleProValLysSerProAspGluProLeu	180
Qy	637	ACTTGGCAGTATGTGATCATGTTTGTCTCGGAATCTGGGGCGTGCAGCACCCTTGGGC	696
Db	181	ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlySerLeuGly	200
Qy	697	TCCCTCGGAAATCGGGAAACAAGGAGAAGAGTCTTCATCAGCTGGTAGCTCCCGA	756
Db	201	SerProGlyAsnArgGluAsnLysGluLysValPheIleSerLeuValGlySerArg	220
Qy	757	GGCCTTGGCTGACAGATPCCAGCGGCCCATCCAGAAAGCTGCGATCTTTATCAGCCAT	816
Db	221	GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis	240
Qy	817	GTGAACCTGGCTCCCTGCTCTCCTCAGTGGGATGGAGATAGGGACACAGATTGTCAA	876
Db	241	ValTyrProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu	260
Qy	877	GTCAATGGCTGCACATCTTCACTAACCTGGATCACAGGAGGCTGTAAATGTCTGAAAAAT	936
Db	261	ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn	280
Qy	937	AGCGGACGCTGACATCTCAATTGTAGCTGCAGCTGCCCGGAGCTGTTCATGACAGAC	996
Db	281	SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp	300
Qy	997	CGGAGCGGCTGCACAGCGCGGAGGTGAGCTGACAGCGCAGGAGCTTCTCATCGAG	1056
Db	301	ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuLeuMetGln	320
Qy	1057	AAGCGGTGCGATGTAGTCCAAACAGATCTCTCCAGGACGACGAGAGATGGAGCGCAA	1116
Db	321	LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGluGlnGluMetGluArgGln	340
Qy	1117	AGGAGAAAAGAAATGCCACAGAGGACGACAGAGGAAAATGAGAGATACCCGGAAGGAGATG	1176
Db	341	ArgArgLysIleIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet	360
Qy	1177	GAACAGATTCTAGCAGGAGGAAGAAGATTAAAGAAGCAATGGGAAGAGACGTGGGGCTCA	1236
Db	361	GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer	380
Qy	1237	AAGAACAGCTACTCTTGGCTTAAACCATCATCTGCTGAGGTACACCCAGTACCCCTTCGC	1296
Db	381	LysGluGlnLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg	400
Qy	1297	AAGCCA-----	1302
Db	401	LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspAspLeuAspGly	420
Qy	1303	-----AAGTGATTTCGGAAATATGAGNAGGCTTTGACCCCTTA	1341
Db	421	GlyThrGluGluGlnGlyGluGln-AspPheArgLysTyrGluGluGlyPheAspProTyr	440
Qy	1342	CTCTATCTTCAACCCAGACAGATCATGGGAAGGATGTCCGGCTCCCTACCCATCAAGAA	1401
Db	440	rSerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLys	460
Qy	1402	GGAGGGATCTCTAGACTGGCCCTGGAAGCGGCTGTGGACTCCCCCATTTGGNAGGTGGT	1461
Db	460	sGluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProIleGlyLysValVa	480
Qy	1462	CGTTTCTGCTGTATAGCGGGGAGCTCTCTGAGCGGCATGGTGGCATTTGTAAAGGGGA	1521
Db	480	IValSerAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAs	500
Qy	1522	CGAGATCATGGCAATCAACGGCAAGATTCTGACAGACTACACCTCGCTGAGGTGACGC	1581
Db	500	pGluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAl	520
Qy	1582	TGCCCTTCAGAAAGCGCTTGAATCAGGCGGGGACTGTGATCGACACTACACCTCGCTGAGGTGACGC	1641

Db 520 aAlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValValalaValcy 540

QY 1642 CCCCCCAAGGAGTATGACGATGAGTACCTTC 1675

Db 540 sProProlGlySerGluLeuThrPhe 551

RESULT 2

Z03_MOUSE

ID Z03_MOUSE STANDARD; PRT; 905 AA.

AC Q90X1.

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona

DE occludens 3 protein) (Tight junction protein 3).

GN TJP3 OR Z03.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALE/C;

RX MEDLINE=20069797; PubMed=10601346;

RA Itoh M., Furuse M., Morita K., Kubota K., Saitou M., Tsukita S.;

RT "Direct binding of three tight junction-associated MAGUKs, ZO-1, ZO-2,

RT and ZO-3, with the COOH termini of claudins.";

RL J. Cell Biol. 147:1351-1363(1999).

CC -|- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1.

CC -|- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.

CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC -|- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL: AF157006; AAF24175.1; -

DR HSP; P31016; IBEF.

DR MGD; MGI:1351650; Tjp3.

DR InterPro; IPR000619; Guanylate_kin.

DR InterPro; IPR01478; PDZ.

DR InterPro; IPR001452; SH3.

DR Pfam; PF00595; PDZ; 3.

DR Pfam; PF00625; Guanylate_kin; 1.

DR SMART; SM00072; GuK; 1.

DR SMART; SM00228; PDZ; 3.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.

DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.

DR PROSITE; PS00002; SH3; FALSE_NEG.

DR PROSITE; PS01016; PDZ; 3.

KW Tight junction; SH3 domain; Repeat; Membrane.

FT DOMAIN 11 93

FT PDZ 1.

FT DOMAIN 187 264

FT PDZ 2.

FT DOMAIN 368 434

FT PDZ 3.

FT DOMAIN 467 540

FT SH3.

FT DOMAIN 654 754

GUANYLATE_KINASE.

SQ SEQUENCE 905 AA; 99324 MW; B787BA1592661FEE CRC64;

Alignment Scores:

Pred. No.: 2,06e-08 Length: 905

Score: 272.00 Matches: 144

Percent Similarity: 38.45% Conservative: 89

Best Local Similarity: 23.76% Mismatches: 203

Query Match: 6.82% Indels: 171

DB: 1 Gaps: 30

US-09-502-945-5 (1-2162) x Z03_MOUSE (1-905)

QY 395 GGCCTCGGCTGAGTGTGCGTGGCTGGAGTTTGGCTGTGGG---CTCTTCATCTCC 441

Db 21 GlyPheGlyIleAlaValSerGlyGlyHisAspArgAlaSerGlySerValValSer 40

QY 442 CACCTCATCAAGCGGCTCAGGCAGACAGCTCGGGCTCCAGGTAGGGAGGAGATGCTC 501

Db 41 AspValValProGlySerProAlaGluGly---ArgLeuArgThrGlyHisPheIleVal 59

QY 502 CGGATCAATGATATTCATCTCCTGTACCCATGAGGAGGTCATCAACCTCATTCGA 561

Db 60 MetValAsnGlyValSerValGluAsnValThrSerAlaPheAlaIleGlnIleLeuLys 79

QY 562 ACC---AAGAAACTGTGTCATCAAAAGTG-----AGACACATCGCCCTGATCCCC 609

Db 80 ThrCysThrLysThrAlaAsnValThrValLysArgProArgValGlnLeuProAla 99

QY 610 GTGAAAAGCTCTCTCTGATGAGCCCTCACCTGGCAGTATGTGGATCATGTTGTCTCGAA 669

Db 100 ThrLysAlaSerProAlaSer-----GlyHisGlnLeuSerAspGlnGluAlaAsp 117

QY 670 TCTGGG----- 675

Db 118 HisGlyArgGlyTyrGluGlyAspSerSerGlySerGlyArgSerTrpGlyGluArg 137

QY 676 -----GGCTGCGAGGAGGCTGGCTCCCTCGGAAATCGGGAAC 717

Db 138 SerArgArgSerArgAlaGlyArgGlyArgValGlySerHisGlyArgArgSer 157

QY 718 -----AAG 720

Db 158 GlyGlySerGluAlaAsnGlyLeuAspLeuValSerGlyTyrLysArgLeuProLys 177

QY 721 GAG-----AAGAGGTCTTCATCAGCCTGTGTAGCTCCCGAGGC 759

Db 178 GlnAspValLeuMetArgProLeuLysSerValLeuValLysArgArgAsnSerGluGlu 197

QY 760 CTGGCTGCGAGCATTTCCACGGCCCATCCAGAGCCTGCATCTTTATCAGCCATGTG 819

Db 198 PheGlyValLysLeuGlySer-----GlnIlePheIleLysHisIle 211

QY 820 AAACCTGGCTCCCTGTCTGTCTGCTG---GAGGTGGATTTGAGATAGGCACCATGTCGAA 876

Db 212 ThrGluSerGlyLeuAlaAlaArgAsnHisGlyLeuGlnGluGlyAspLeuIleLeuGln 231

QY 877 GTCATGGCTCGACTTCTCTAACCTGGATCACAAGGAGGCTGTAATGTGCTGAAAAAT 936

Db 232 IleAsnGlyValSerSerAlaAsnLeuSerLeuSerAspThrArgArgLeuIleGluLys 251

QY 937 AGCCGC---AGCCTGACCATCTCCATTTAGCTGCAGCTGGCGGGAGCTGTTTCATGACA 993

Db 252 SerGluGlyGluLeuThrLeuLeuValLeuArgAspSerGlyGlnPheLeuVal---AsnI 271

QY 994 GACCGGAGCGGCTGGCAGAGCGCGCGCTGAGCTGCAGCGGAGGAGCTTCTCATG 1053

Db 271 leProProAla-----ValSerAspSerAspSerSerLeuM 283

QY 1054 CAGAGCGGCTGGCGAT-----GGAGTCCAACAAGATCTCCAGGAGCAGCAGGAG 1104

Db 283 etGluAspIleSerAspLeuThrSerGluLeuSerGlnAlaProProSerHisValProp 303

QY 1105 ATGAGCGGCAAGGAGAGAAAGAAATGCCCAAGGAGCAGCAGGAGAAATGAGATAC 1164

Db 303 roProProLeuLysGlyGlnArgSer-----p 312

QY 1165 CGGAGGAGATGCAACAGATTGTAG-----AGGAGAAAGAGAA 1202

Db 312 roGluAspSerGlnThrAspSerProValGluThrProGlnProArgArgGluArgS 332

QY 1203 GTTTAAGAACAATGGGAAGAAGACTGGGCTCAAAGGAACAGCTACTCTTGCCTAAAC 1262

Db 332 erValAsnSer-----ArgAlaIleAlaGluProGluSer-----p 344

QY 1263 CATCACTGCTGAGTACACCCAGTACCCCTTCGCAAGCAAAAGTGATTCCTCGGAATATG 1322
 Db 344 roGlyGluSerArgTyrAspIleTyrArgValProSerArgGlnSerLeuGlu-----A 362
 QY 1323 AGGAAGGCTTGGACCCCTACTCTATGTTTCACCCAGACAGATCATGCGGAAGATGTC 1382
 Db 362 spArgGlyTyrSerPro-----AspThrA 370
 QY 1383 GGCTCTACGATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAAGGGGGTGGACT 1442
 Db 370 rgValSerPheProGlySerGlyAlaSerIleGlyLeuArgLeuAlaGlyGlyAsnAsp- 389
 QY 1443 CCCCCATTGGGAAGTGTCTGCTGTATGAGCGGGAGCTGCTGAGCGGATG 1502
 Db 390 -----ValGly---IlePheValSerGlyVal---GlnAlaGlySerProAlaAspGlyG 406
 QY 1503 GTGCAATTGTGAAGGACGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACA 1562
 Db 406 lngIleGlnGluGlyAspGluIleLeuGlnValAsnGlyMetProPheArgAsnLeuT 426
 QY 1563 CCCTGGCTGAGGCTGACCTCCCTGCGAAGAGGCTGGAATCAGCGGGGACTGGATCG 1622
 Db 426 hrArgGluGluAla-----ValG 432
 QY 1623 ACCTGTGCTGCTGCTGCTGCCCCCAAGAGGATGATGACGATGAGTACCTTCTTGCTGA 1682
 Db 432 lnPheLeuLeuGlyLeu---ProProGlyGlu-----AspMetGluLeuValThrG 448
 QY 1683 AGTCCAAAGGGAACCAACAAATTCACGGGTAGGAACACAGTGAAGTCCGGCCCACTCG 1742
 Db 448 lnSerLysThrGlyHisSerLeuArgArgTyrSerGlnSerArgVal-----463
 QY 1743 TGAACACAAAGCCTGGACCCCTGAGAGAGGCCATGACACACACAGATGGCATC 1802
 Db 464 -----GlyAspSerPheTyrIleArgThrHis- 472
 QY 1803 CTTGGGACCTGAATCATCACCCAGGATCTCAAACTCCCTTGGCCCTGAACACAGGCC 1862
 Db 473 --PheGluLeuGluProSerPro-----ProTyrGlyLeuGlyPheThrArgGlyA 489
 QY 1863 AGATA-----AGAACAGCTCGGGCCACTTTTGTGAAG 1895
 Db 489 spValPheHisValValAspThrLeuTyrProGlySerGlyProGlyHisGlyHisSers 509
 QY 1896 GCCAATGTGGAGAAAGGAGGACGAGCAGCGCTTGGGAGAGATCTCAAGGATCCAGACT 1955
 Db 509 exHis---GlyGlyLeuTrpLeuAlaArgMetGlyArgAspLeuArgGluGlnGluA 528
 QY 1956 CTCATTCTCTTCTTCT 1969
 Db 528 rgGlyValIlePro 532
 RESULT 3
 Z03_HUMAN
 ID Z03_HUMAN STANDARD; PRT; 933 AA.
 AC O95049.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tight junction protein 20-3 (Zonula occludens 3 protein) (Zona
 DE occludens 3 protein) (Tight junction protein 3).
 GN TJP3 OR Z03.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,
 RA Burkhardt-Schultz K., Brower A., Gordon L., Dias J., Ramirez M.,
 RA Stillwagen S., Phan H., Velasco N., Do L., Regala W., Terry A.,
 RA Garnes J., Danganan L., Erler A., Christensen M., Georgescu A.,

RA Avila J., Liu S., Attix C., Andreise T., Trankheim M.,
 RA Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R.,
 RA Thomas P., Quan G., Krommiller B., Arellano A., Sanders C., Ow D.,
 RA Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.,
 RT "sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
 RT serine protease gene cluster."
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBUNIT: INTERACTS WITH OCCUDLIN, CLAUDINS AND ZO-1 (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AC005954; AAC72274.1; ALT_INIT.
 DR HSSP; P31016; IBFE.
 DR Genew; HGNC:11829; TJP3.
 DR InterPro; IPR000619; Guanylate_kin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00595; PDZ; 3
 DR Pfam; PF00625; Guanylate_kin; 1.
 DR SMART; SM0072; Gukc; 1.
 DR SMART; SM00228; PDZ; 3.
 DR SMART; SM0326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
 DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS50002; SH3; 1.
 DR PROSITE; PS50106; PDZ; 3.
 DR KW Tight junction; SH3 domain; Repeat; Membrane.
 FT DOMAIN 11 93
 FT DOMAIN 195 272 PDZ 1.
 FT DOMAIN 394 460 PDZ 2.
 FT DOMAIN 489 563 SH3.
 FT DOMAIN 675 775 GUANYLATE KINASE.
 SQ SEQUENCE 933 AA; 102800 MW; 680298CFD0615B47 CRC64;
 Alignment Scores:
 Pred. No.: 2,36e-08 Length: 933
 Score: 271.00 Matches: 144
 Percent Similarity: 37.84% Conservative: 91
 Best Local Similarity: 23.19% Mismatches: 207
 Query Match: 6.79% Indels: 179
 DB: 1 Gaps: 30
 US-09-502-945-5 (1-2162) x Z03_HUMAN (1-933)
 QY 385 GGCCTCGCCCTGAGTGTGCTGGTGGCTGGAGTTT---GGCTGTGGCTTTTCATCTCC 441
 Db 21 GlyPheGlyIleAlaIleSerGlyArgAspArgProGlyGlySerMetValValSer 40
 QY 442 CACCTCATCAAGGGGTGAGGACAGACAGCGTCCAGGTAGGAGGACGAGATCGTC 501
 Db 41 AspValValProGlyGlyProAlaGluGly---ArgLeuGlnThrGlyAspHisIleVal 59
 QY 502 CGGATCAATGATATTCATCTCTCTGTATCCATGAGGAGGTTCATCAACCTCATTCGA 561
 Db 60 MetValAsnGlyValSerMetGluAsnAlaThrSerAlaPheAlaIleGlnIleLeuLys 79
 QY 562 ACC---AAGAAACTGTCTCCATCAAGTG-----AGACACATCGGCTGATCCCC 609
 Db 80 ThrCysThrLysMetAlaAsnIleThrValLysArgProArgGileHisLeuProAla 99
 QY 610 GTGAAAGCTCTCTCT---GATGAGCCCTC 636

```
Db 100 ThrLysAlaSerProSerSerProGlyArgGlnAspSerAspGluAspAspGlyProGln 119
QY 637 ACTTGGCAGTATGTGGATCAG-----TTTGTGTCGGAATCTGGG 675
Db 120 ArgValGluGluValaspGlnGlyArgGlyTyraSpGlyaspSerSerGlySerGly 139
QY 676 -----GGCGTGCAGGAGCAGCCTGGGGCTCC 699
Db 140 ArgSerTrpAspGluArgSerArgProArgProGlyArgGlyArgAlaGlySer 159
QY 700 CCTGGAATCGGAAACAAGGAGAAGGTTCTCATCAGCCTGTAGCTCCGAGGC 759
Db 160 HisGlyArgArgSerProGlyGly-----GlySerGluAla 171
QY 760 CTTGGTGCAGCATTTCCAGCGC-----CCCATCCAG-----AAG 795
Db 172 AsnGlyLeuAlaLeuValSerGlyPheLysArgLeuProArgGlnaspValGlnMetLys 191
QY 796 CCH-----798
Db 192 ProValLysSerValLeuValLysArgArgAspSerGluGluPheGlyValLysLeuGly 211
QY 799 ---GGCATCTTTATCAGCCATGTGAACCTGGCTCCTGCTCTCT---GAGGTGGATTC 852
Db 212 SerGlnIlePheIleLysHisIleThrAspSerGlyLeuAlaAlaArgHisArgGlyLeu 231
QY 853 GAGATAGGGAGCAGATTCGGAAGTCAATGGCGTGGACTTCTTAACCTGGATCAACAAG 912
Db 232 GlnGluGlyAspLeuIleAsnGlyValSerSerGlnAsnLeuSerLeuAsn 251
QY 913 GAGCGTGTAAATGCTGTAATAATACCGCAGCGCTGACCATCTCCATTCAGTCAGCT 972
Db 252 AspThrArgLeuLeuGluGlySerGluGly---LysLeuSerLeuLeuValLeuArg 270
QY 973 GGCGCGGAGCTGTTCATGACAGACCGGAGCGCGCTGGCAGCGCGGAGCTGAGCTG 1032
Db 271 AspArgGlyClnPheLeuValAsnIleProProAlaValSerAspSerAspSerPro 290
QY 1033 CAGCGCGAGGAGCTTCTCATGAGAAAGCGGCTGGCGATGAGTCCAAAGATCCTCCAG 1092
Db 291 LeuGluGluGlyValThrMetAlaaspGluMetSerSerProAlaaspIleSerAsp 310
QY 1093 GAGCAGCAGGATGGAGCGG-----1113
Db 311 LeuAlaSerGluLeuSerGlnAlaProProSerHisIleProProProArgHisAla 330
QY 1114 CAAAGGAGAAAGAAATTCGCCAG---ANGGAGCAGAGAGAAATCAGAGATACCGAAG 1170
Db 331 GlnArgSerProGluAlaSerGlnThrAspSerProValGluSerProArgLeuArgArg 350
QY 1171 GAGATG-----GAACAGATTGTAGAGGAGGAGAGAGTTTAAAGAACCAATGGGA 1221
Db 351 GluSerSerValaspSerArgThrIleSerGluProAspGluGlnArgSerGlu-----368
QY 1222 GAAGACTGGGCTCAAGGAGACAGTACTCTTCCTTAAACCATCATCTGCTGAGGTACAC 1281
Db 369 -----LeuProArgGluSerSerTyraSpIleTyra 378
QY 1282 CCAGTACCCCTTCGCAAGCCAAAGTCATTCTCCGGAATATGAGGAGCGTTTGGACCCCTA 1341
Db 379 ArgVal-Pro-----SerSerGlnSer-----MetGluAspArgGlyTyraSerPro---393
QY 1342 CTCATGTTCACCCAGCAGCATCATGGGAGAGATGTCGGCTCCCTACGATCAAGAA 1401
Db 394 -----AspThrArgValValArgPheLeuLys 402
QY 1402 GGAGGATCTTTAGACCTGGCCCTGGAAGCGGTGTGGACTCCCTCCATTGGGAAGTGT 1461
Db 402 scelyLysSerIleGlyLeuArgLeuAlaGlyLysAsnasp-----ValGly---IlePh 419
QY 1462 CGTTTCTGCTGTATGACGGGGAGCTGCTGAGCGGCATGTGTGATGTGAAGAGGGA 1521
Db 419 eValSerGlyVal---GlnAlaGlySerProAlaAspGlyGlnGlyIleGlnGluGlyAs 438
```

```
QY 1522 CGAGATCATCGGCATCAACGGCAAGATTGTGACAGACTACACCTGGCTGAGGCTGACGC 1581
Db 438 pGlnIleLeuGlnValAsnAspValProPheGlnAsnLeuThrArgGluGluAla-----456
QY 1582 TGCCCTGCAGAGGCGCTGGAATCAGGGCGGACTGATCGACCTTGCTGTGCTCCCTCTG 1641
Db 457 -----ValGlnPheLeuLeuGlyLeuPr 464
QY 1642 CCCCCCAAAGGAGTATGAC-----GATGAGCTGACCTTCTTGTCT 1680
Db 464 oProGlyGluGluMetGluLeuValThrGlnArgLysGlnAspIlePheTrpLysMetVa 484
QY 1681 GAAGTCCAAAGGGGAAACCAATTCACCGGTAGGAACAGTGAAGTCCGGCCCAACCT 1740
Db 484 IginSerArgValGlyAspSerPheTyraIleArgThrHisPheGluLeuGluPro-----502
QY 1741 COTGAACAAAGCCTCGGACCAAGCCTT-----GAGAGAGCGGCACATCACACACC--1792
Db 503 -----SerProSerGlyLeuGlyPheThrArgGlyAspValPheHisValle 519
QY 1793 ---AGATGGCATCTTGGGACCTGAATCTATCACCAGGAATCTCAAACTCCCTTTGGC 1848
Db 519 uAspThrLeuHisPro-----524
QY 1849 COTGAACAGGCGCATATAAGGAACAGCTCGGGCCACTTTTGAAGGCCAATGTGGAGG 1908
Db 525 ---GlyProGlyGlnSerHisAlaArgGlyGlyHisTrpLeu-----537
QY 1909 AAGGAGCAGCAGCCGCTTGGGAGAGATCTCAAGATCCAGACTCTCATTCCTTTCC 1968
Db 538 -----AlaValArgMetGlyArgAspLeuArgGluGlnGluArgGlyIlelepr 554
QY 1969 T 1969
Db 554 o 554
RESULT 4
203_CANFA
ID Z03_CANFA STANDARD; PRT; 898 AA.
AC O62683;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-3 (zonula occludens 3 protein) (Zona
DE occludens 3 protein) (tight junction protein 3).
GN TJP3 OR Z03.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=98198478; PubMed=9531559;
RA Haskins J., Gu L., Wittchen E.S., Hubbard J., Stevenson B.R.;
RT "ZO-3, a novel member of the MAGUK protein family found at the tight
RT junction, interacts with ZO-1 and occludin.";
RL J. Cell Biol. 141:199-208(1998).
CC -|- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1.
CC -|- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF023617; AAC39177.1; -.
```

DR	HSP: P31016; 18PF	
DR	InterPro: IPR000619; Guanylate_kin.	
DR	InterPro: IPR001478; PDZ.	
DR	InterPro: IPR001452; SH3.	
DR	Pfam: PF00595; PDZ; 3.	
DR	Pfam: PF00625; Guanylate_kin; 1.	
DR	SMART: SM00072; GuK; 1.	
DR	SMART: SM00228; PDZ; 3.	
DR	SMART: SM00326; SH3; 1.	
DR	PROSITE: PS00856; GUANYLATE_KINASE_1; FALSE_NEG	
DR	PROSITE: PS00502; GUANYLATE_KINASE_2; 1.	
DR	PROSITE: PS00106; PDZ; 3.	
DR	PROSITE: PS00002; SH3; 1.	
KW	Tight junction; SH3 domain; Repeat; Membrane.	
FT	DOMAIN 11 93	PDZ 1.
FT	DOMAIN 187 264	PDZ 2.
FT	DOMAIN 369 435	PDZ 3.
FT	DOMAIN 464 538	SH3.
FT	DOMAIN 650 750	GUANYLATE_KINASE.
SQ	SEQUENCE 898 AA; 98414 MW; 8091D6132D89F15D	

Alignment Scores:	
Pred. No.:	1,828-07
Score:	285.50
Percent Similarity:	38.26%
Best Local Similarity:	24.30%
Query Match:	6.41%
DB:	1
Length:	898
Matches:	148
Conservative:	85
Mismatches:	197
Indels:	180
Gaps:	31

US-09-502-945-5 (1-2162) x 203_CANFA (1-898)

[illegible]

Db	214	SerGlyLeuAlaAlaAsnArgGlyLeuGlnGluGlyAspLeuIleLeuGlnIleAsn	233
Qy	883	GGCGTCGACTTCTTAACCTGGATCACAAGGAGCGTGTAAATGTGCTGAAAAATAGCCGC	942
Db	234	GlyValSerSerGluAsnLeuSerLeuSerAspThrArgLeuLeuGluLysSerGlu	253
Qy	943	--AGCCTGACCATCTCCATTCTAGCTCAGCTGCGCGGAGCTGTCATGACAGACACGG	999
Db	254	GlyLysLeuThrLeuLeuValLeuAlaArgAspArgGlyGlnPheLeuVal--AsnIleProp	273
Qy	1000	GAGCGCTGCGCAGAGCGCGCGCTGAGCT-----GCAGCGCAGGAG	1044
Db	273	roAlaValSerAspSerAspSerAspSerPheLeuAspAspIleSerAlaLeuGlyS	293
Qy	1045	CTTCTCATGACAGAGCGGCTGGAGTGAGTCCACAAGATCTCTCAGGAGCAGCAGGAG	1104
Db	293	erGluLeuSerGlnAlaValProSerHisValProProPheHisAla-----	310
Qy	1105	ATGGAGCGCAAGGAGAAACAATTCGCCAGAGGCGCAGAGGAATGAGAGATAC	1164
Db	311	-----GlnArgSerLeu-----	314
Qy	1165	CGGAAGGAGATGGAACAGATTAG-----AGGAGGAAGAGAAGTT	1205
Db	315	--AspSerAspGlyThrAspSerProArgAspSerProLeuArgArgGluAsnSerL	334
Qy	1206	TAGAGACCAATGGAGAGACAGATGGGGTCAAGAGAACAGCTACTCTTGCTTAAACCAT	1265
Db	334	eu-----AspSerAspThrIleSerGluPro	343
Qy	1266	CAGTGTGAGTACACCCAGTACCCCTTC-----GCAAGCCAAAGTGATTTC	1313
Db	343	spAlaProArgHisSerSerTyrAspIleTyrArgValProSerSerGlnSer-----	360
Qy	1314	GGAAATATGAGGAGGCTTGTACCCCTACTCTTATCTCACCAGAGCATATGCGGA	1373
Db	361	--AlaGluAspArgGlyTyrSerPro-----	368
Qy	1374	AGGATGTCGGCTCTACCGATCAAGAAGGAGGATCTTAGACCTGCGCTTGAAGCGC	1433
Db	369	--AspSerArgValValArgPheHisLysGlyThrIleGlyLeuArgLeuAlaGly	388
Qy	1434	GTGTGGACTCCCCATTTGGAGGTGGTTCGTCTGTGTATGAGCGGGAGCTGCTG	1493
Db	388	lyAsnAsp-----ValGly--IlePheValSerGlyValGlnGlySerProAla	405
Qy	1494	AGCGGCATGTGGCATTTGAAAGGGAGCAGATCATGGCAATCAACGGCAAGATTGTA	1553
Db	405	spGlyGln--GlyIleGlnGluGlyAspGlnIleLeuGlnValAsnAspValProPheA	424
Qy	1554	CAGACTACACCTGGCTGAGGTGAGCTGCGCTGCCAGAGGCCTGGAAATCAGGCGGGG	1613
Db	424	rgAsnLeuThrArgGluGluAla-----	431
Qy	1614	ACTGGATCGACCTGTGTGTTCGCTGCCCGCCCAAGGAGTATGACGATGAGCTGACCT	1673
Db	432	-----ValGlnPheLeuValAlaLeuProGlyGluGluValGlu-----	445
Qy	1674	TCCTTGCTGAATCCAAAGGGAAACCAAAATTCACGCGTTAGGAAACAGTAGTCCGGC	1733
Db	446	-----LeuValThrGlnArg-----AsnGluAspIlePheA	456
Qy	1734	CCCACCTCGTGAACACAAGCCTCGGACCGCTTGAGAGAGGCCACATGACACACCA	1793
Db	456	rgLysMetValGlnSerArgVal-----GlyAspSerPheTyrIleA	470
Qy	1794	GATGGCATCTTGGGACCTGAATCTATCACCAGGAATCTCAACTCCTTTGGCCCTGA	1853
Db	470	rgThrHis--PheGluLeuAlaSerPro-----ProSerGlyLeu-	483
Qy	1854	ACCAGGCCAGATAAGAACAGCTCGGGCCACTTTTGAAGGCCCAATGTGGGAAGG	1913
Db	484	-----GlyPheThrArgGlyAspValPheHisValLeuAspThrLeuCysProGlyPro	502


```
QY 968 CAGCTGGCGGAGCTGTTTCATGACACAGCGGAGCGCTGCAGAGCGCGGCGAGCTG 1027
DB 329 yProGly-----SerIleHisLeuAsnA 337
QY 1028 AGCTGACGCGGAGGAGCTTCTCATGCAGAACGGCTGGCGATGGAGTCCCAACAGATPC 1087
DB 337 spMetTyraIaProProAspTyraIaSerThrPheThrAlaLeuAlaAspAsnHisIle- 356
QY 1088 TCC---AGAGCAGCAGGAGATGGAGCGGCAAGAGGAGAGAAATTCGCCAGAGGCGAG 1144
DB 357 SerHisAsnSerSerLeuGlyTyraIaValGluSerLysVal----- 372
QY 1145 CAGAGGAAATCAGAGATACCGGAGGAGATGGAACAGATTCTACAGGAGGAAGAAGT 1204
DB 372 ----- 372
QY 1205 TTAAGAAGCAATGGGAAGAGACTGGGCTCAAAGAACAGACTACTCTTGCCTAAACCA 1264
DB 372 ----- 372
QY 1265 TCACTGCTGAGGTACACCCAGTACCCCTCGCAAGCCAAAGTATTCCGGAAATATGAG 1324
DB 373 -----ThrTyraIaProProGlnValProProThrArgTyraIaSer 386
QY 1325 GAAGGCTTGACCCCTACTCTATGTTTACCCAGCAGATCATGGGGAAGGATGTCGGG 1384
DB 387 Pro-----IleProArgHisMetLeuAlaGluAlaGluAspPheThrArgGluProArgLys 404
QY 1385 CTCCTACGATCAAGAGGAGGATCTTACCTGAGCCTGGAGCGGCTGGAGCTGACTCC 1444
DB 405 IleIleLeuHisIleGlySerThrGlyLeuGlyPheAsnIleValGlyGlyGluAspGly 424
QY 1445 OCCATTGGGAGGTGCTCTCTCTCTGCTGTATGAGCGGGGAGCTGCTGAGCGGATGGT 1504
DB 425 Glu-----GlyIlePheValSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGly 442
QY 1505 GGCATTGTGAAGGGGAGCAGATCATGGCAATCAAGCGCAAGATTGTGACAGCTACAC 1564
DB 443 GluLeuArgArgGlyAspArgIleLeuSerValAsnGlyValAsnLeuArgAsnAlaThr 462
QY 1565 CTGGCTGAGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1624
DB 463 HisGluGlnAlaAlaAlaLeuLysArgAla-----GlyGlnSerVal---- 477
QY 1625 CRTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1684
DB 478 ThrIleValAlaGlnTyrArgProGluGluTyr----- 488
QY 1685 TCCAAAGGGGAAACCAAAATTCACGCTTAGGAACAGTCAGCTCGCGGCCACCTCGTG 1744
DB 489 SerArgPheGluSerLysIleHis-----AspLeuArgGluGlnMetMet 503
QY 1745 AACACAAAGCCT-----CGGACAGCCTTGAGAGA 1774
DB 504 AsnSerMetSerSerGlySerGlySerLeuArgThrSerGluLysArg 520
RESULT 6
DLG3_RAT STANDARD; PRT; 849 AA.
AC Q62936; P70547;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic protein SAP102 (Synapse-associated protein 102) (PSD-
DE 95/SAP90 related protein 1) (Discs, large homolog 3).
GN DLG3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE=Brain;
```

```
RX MEDLINE-96374358; PubMed-8780649;
RA Mueller B.M., Kistner U., Kindler S., Chung W.J., Kuhlendahl S.,
RA Fenster S.D., Lau L.-F., Veh R.W., Hugenir R.L., Gundelfinger E.D.,
RA Garner C.C.;
RT "SAP102, a novel postsynaptic protein that interacts with NMDA
RT receptor complexes in vivo.";
RL Neuron 17:255-265(1996).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RA Irie M., Hata Y., Takai Y.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC SUBUNIT NR2B.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U50147; AAA93031.1; -.
DR EMBL; U53367; AAB48561.1; -.
DR HSPF; Q12959; IPDR.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00072; Gukc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
DR SH3 domain; Repeat; Alternative splicing.
KW SH3 domain; Repeat; Alternative splicing.
FT DOMAIN 149 235 PDZ 1.
FT DOMAIN 244 330 PDZ 2.
FT DOMAIN 404 484 PDZ 3.
FT DOMAIN 519 589 SH3.
FT DOMAIN 659 849 GUANYLATE_KINASE.
FT VARSPLOC 627 640 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 849 AA; 93539 MW; 34DA9C46C7BB96DB CRC64;
Alignment Scores:
Pred. No.: 3.99e-07 Length: 849
Score: 249.50 Matches: 139
Percent Similarity: 36.59% Conservative: 82
Best Local Similarity: 23.01% Mismatches: 191
Query Match: 6.25% Indels: 192
DB: 1 Gaps: 26
US-09-502-945-5 (1-2162) x DLG3_RAT (1-849)
QY 29 CCAGCTCTCTGCGAGCGCGGACCCGGAAGGAGCGTGTGCAACGACGCGCTGGACCTG 88
DB 89 ProValProGlyLysAsnThrProLysLeuAsnGlySerGlyProSer----- 104
QY 89 GCCCAGCCATGACCGGACGAGGATTCGGCAGATTCGGCAGATGAGTTCGATG 148
DB 105 -----TriProGlu----- 108
QY 149 AAAATGATGCAGAGAAGGACTATCTATGATGCTGCGGAATGTACACACGACCATGG 208
DB 109 -----CysThrCysThrAsnArgAspTrp 116
```


QY 209 ACGTGCCGCTGCTCGTGGGAGACCTCAAGCTGGTTCATCAATGAACCCAGCCGCTGCTC 268
 Db 117 -----TyrGluGlnAlaSerProAlaPro---Leu 125
 QY 269 TGTATTGATCCATTCGGCGGTGATCCCACTGAACACAGCGTGAATATGATCAGCTGA 328
 Db 126 LeuValasnPro-GluAlaLeuGluProSerLeuSerValasnGly-SerAspGlyMetp 145
 QY 329 CCCCCCGGCTCCAGGAGGCTGAAGGAGTCCGCTGCTGGACCGTGTGACCCCGGAAGCC 388
 Db 145 he-----LysTyrGluGluIleValLeuGluArgGlyAsn---SerGlyL 159
 QY 389 TCGGCTGAGTGTGGCTGGCTGGAG-----TTTGGCTGTGGCTCT 433
 Db 159 euGlyPheSerIleAlaGlyIleAspAsnProHisValProAspAspProGlyIleP 179
 QY 434 TCATCTCCACCTCATCAAGGCGGTGAGGAGAGCGGTGGG---CTCAGGTAGGG 490
 Db 179 heIleThrLysIleIleProGlyAlaAlaAlaMetAspGlyArgLeuGlyValasnA 199
 QY 491 ACGAGATCGTCCGATCAATGATATCCATCTCTCTCTGCTGACCATGAGGAGGTATCA 550
 Db 199 spCysValLeuArgValasnGluValaspValSerGluValHisSerArgAlaValG 219
 QY 551 ACCTCATTCGAACCAAGAAACTGTGTCCATCAAGTGAAGACATCGGCTGATCCCG 610
 Db 219 luAlaLeuLysGluAlaGlyProVal-----ValArg-----LeuValV 232
 QY 611 TGAAGATCTCTCTGATGAGCCCTCACTTGGCGATGATGTGGATCAGTTGTGTGCGAAT 670
 Db 232 alaArgArgGlnProProGluThrIleMetGluValasn----- 246
 QY 671 CTGGGGCGTGGAGCGAGCTGGCTCCCTCGGAATCGGAAACAGGAGAGGAGG 730
 Db 246 ----- 246
 QY 731 TCTTCATCAGCTGTAGTCCGAGGCGCTTGGTGCAGCATTTCCAGCGGCCCATCC 790
 Db 247 -----LeuLeuLysGlyProLysGlyLeuGlyPheSerIleAlaGlyIleGlyAla 264
 QY 791 AGAAG-----CCTGGC-----ATCTTATCAGCATGTGAAACCTGGCTCGCTGT 835
 Db 264 snGlnHisIleProGlyAspAsnSerIleTyrIleThrLysIleIleGlyGlyAlaA 284
 QY 836 CTGCTGAGGTGGGA---TTGAGATAGGGACAGATGTGCGAAGTCAATGCGCTGACT 892
 Db 284 laGlnLysAspGlyArgLeuGlnIleGlyaspArgLeuLeuAlaValasnThrAsnL 304
 QY 893 TCTTAACCTGGATCAAGAGGCTGTAATGTGCTGAAATAATAGCCGAGCCTGACCA 952
 Db 304 euGlnAspValArgHisGluAlaValAlaSerLeuLysAsnThrSerAspMetValt 324
 QY 953 TCTCCATTGATGTCAGCTGGCGGAGCTGTCATGACAGACCGGAGCGCTGCGAG 1012
 Db 324 yrLeuLysValAlaLysProGly---SerLeuHisLeuAsnAsp----- 337
 QY 1013 AGCGCGGAGCTGAGCTGCGAGCGGAGGAGCTTCTCATGCAAGCGGTGCGATGG 1072
 Db 338 -----MetTyrAlaProProAspTyrAlaSerThrPheThrAlaLeuA 352
 QY 1073 AGTCCACAGATCTCTC---AGGAGCAGGAGGATGGAGCGCAAGGAGGAGGAAA 1129
 Db 352 laAspAsnHisIle-SerHisAsnSerSerLeuGlyIleValGluSerLys 371
 QY 1130 TTGCCAGAGGCGAGCAGAGAAATGAGATACCGGAGGAGATGGAACAGATTGTAG 1189
 Db 372 Val----- 372
 QY 1190 AGGAGGAAGAGAGTTTAAAGCAATGGGAAGAAGACTGGGGCTCAAGGAGACAGCTAC 1249
 Db 372 ----- 372

QY 1250 TCTTGCCCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCAAGTGAT 1309
 Db 373 -----ThrTyrProAlaProProGlnValPro 381
 QY 1310 TTCGGAAATATGAGGAAGGCTTTGACCCCTACTCTATTCACCCCGACAGCAGATCATG 1369
 Db 382 ProThrArgTyrSerPro-----IleProArgHisMetLeuAlaGluAspPheThr 399
 QY 1370 GGAAGAGATGTCGGCTCTACGATCAAGAAGGAGGATCTTTAGACCTGGCCCTGAA 1429
 Db 400 ArgGluProArgLysIleIleLeuHisLysGlySerThrGlyLeuGlyPheAsnIleVal 419
 QY 1430 GGCGGTGTTGGACTCCCATTTGGGAAGGTGGTCTCTCTGCTGTGTATGAGCGGGAGCT 1489
 Db 420 GlyGlyGluaspGlyGlu-----GlyIlePheValSerPheIleLeuAlaGlyGlyPro 437
 QY 1490 GCTGAGGCGGATGTGGCATTTGTGAAGGGGAGGAGATGTCGCAATCAACGGCAAGATT 1549
 Db 438 AlaAspLeuSerGlyGluLeuArgArgGlyAspArgIleLeuSerValasnGlyValasn 457
 QY 1550 GTGACAGATACACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1609
 Db 458 LeuArgAsnAlaThrHisGluGlnAlaAlaAlaAlaLeuLysArgAla----- 473
 QY 1610 GGGGACTGGATCGACCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1669
 Db 474 GlyGlnSerVal---ThrIleValAlaGlnTyrArgProGluGluTyr----- 488
 QY 1670 ACCTTCTTGTGAAGTCCAAAGGGGAAACCAATTCACGCGTTAGGAACAGTGAGCTC 1729
 Db 489 -----SerArgPheGluSerLysIleHis-----AspLeu 498
 QY 1730 CGGCGCCACCTCGTGAACACAAAGCCCT-----CGGACACAGCTT 1768
 Db 499 ArgGluGlnMetMetAsnSerSerMetSerGlySerGlySerLeuArgThrSerGlu 518
 QY 1769 GAGAGA 1774
 Db 519 LysArg 520
 RESULT 7
 ZOI_MOUSE STANDARD; PRT; 1745 AA.
 ID ZOI_MOUSE AC P39447;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tight junction protein ZO-1 (Zonula occludens 1 protein) (Zona
 occludens 1 protein) (Tight junction protein 1).
 GN TP1 OR ZO1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=93352986; PubMed=8486731;
 RA Itoh W., Nagafuchi A., Yonemura S., Yasuda-Kitani T., Tsukita S.,
 RA Tsukita S.;
 RT "The 220-kD protein colocalizing with cadherins in non-epithelial
 RT cells is identical to ZO-1, a tight junction-associated protein in
 RT epithelial cells: cDNA cloning and immunoelectron microscopy.";
 RL J. Cell Biol. 121:491-502(1993).
 CC -!- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL
 CC REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY
 CC HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN
 CC MIGHT BE INVOLVED IN STABILIZING JUNCTIONS.
 CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-3.
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.
 CC MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT
 CC OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

CC EMBL: D14340; BAA03274.1; -
 DR PIR: A46431; A46431.
 DR HSP: P31016; IBE9.
 DR MGD: MGI:98759; Tjpl.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR000906; ZU5.
 DR Pfam: PF00018; SH3; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00791; ZU5; 1.
 DR SMART: SM00072; GuK; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR SMART: SM00218; ZU5; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
 DR PROSITE: PS00502; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS0106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 DR Tight junction; SH3 domain; Repeat; Membrane.
 KW DOMAIN 23 110 186 264
 FT DOMAIN PDZ 1.
 FT DOMAIN PDZ 2.
 FT DOMAIN PDZ 3.
 FT DOMAIN SH3.
 FT DOMAIN 516 584
 FT DOMAIN 644 794
 FT DOMAIN 1242 1247
 FT DOMAIN POLY-PRO.
 FT DOMAIN 1424 1430
 SQ SEQUENCE 1745 AA; 194710 MW; C3DA2C0A9F411F66 CRC64;

Alignment Scores:

Pred. No.: 2,7e-06 Length: 1745
 Score: 236.00 Matches: 146
 Percent Similarity: 37.63% Conservative: 102
 Best Local Similarity: 22.15% Mismatches: 218
 Query Match: 5.92% Indels: 193
 DB: 1 Gaps: 32

US-09-502-945-5 (1-2162) x ZOL_MOUSE (1-1745)

Qy 358 GTGCGCTGGACCGTCTGCACCCGAGCGCTGGCTGAGTGTGCGTGGCTGGAG 417
 Db 24 ValThrLeuHisArgAlaProGlyPheGlyPheGlyIleAlaIleSerGlyGlyArgasp 43
 Qy 418 -----TTTGGCTGGG-----CTCTCATCTCCACCTCATCAAGGCGGT 459
 Db 44 AsnProHisPheGlnSerGlyGluThrSerIleValIleSerAspValLeuLysGlyGly 63
 Qy 460 CAGGACAGACGCGTCCAGGTGAGGACGAGATCGTCCGAGTCATGATATTC 519
 Db 64 ProAlaGluGly---GlnLeuGlnGluAsnAspArgValAlaMetValAsnGlyValSer 82
 Qy 520 ATCTCTCTGTACCATTGAGGAGTCAATCAAC----- 552
 Db 83 MetAspAsnValIleHisAlaPheAlaValGlnGlnLeuArgLysSerGlyLysAsnAla 102
 Qy 553 -----CTCATTCGACCAAGAACTGTGTCATCAAGTCAGACACATCGGCTGATC 606
 Db 103 LysIleThrIleArgArgLysLysValGlnIleProValSerHis----- 118
 Qy 607 CCCGTGAAAGCTCTCCTGAT---GAGCCCTCACTTGGCAGTATGTGATCATGTTGTG 663

Db 119 -----ProaspProGluProValSerAspAsnGluAspSerTyrAsp 133
 Qy 664 TCGGAATCTGGGGCGCTGGAGGACCTCGCTGGAAATCGGAAACAGAGAG 723
 Db 134 GluGluValHisAspProAlaGlyArgGlyAlaLeuAlaAsnArgSerGluLys 153
 Qy 723 ----- 723
 Db 154 SerTrpAlaArgAspArgSerAlaSerArgGluArgSerLeuSerProArgSerAspArg 173
 Qy 724 -----AAGAGGTCTTCATCAGCTGCTGAGTCCCGA 756
 Db 174 ArgSerValAlaSerSerGlnProAlaLysProThrLysValThrLeuValLysSerArg 193
 Qy 757 -----GGCCTTGGCTGCAGCATTTCCAGCGCGCCCATCCAGAACCTGGC 801
 Db 194 LysAsnGluGluTyrGlyLeu-----ArgProAla 203
 Qy 802 -----ATCTTTATCAGCCCATGTGAACCTGGCTCCTGTCTGTCTGAGTGGGA---TTG 852
 Db 204 SerHisIlePheValLysGluIleSerGlnAspSerLeuAlaAlaArgAspGlyAspIle 223
 Qy 853 GAGATAGGGACGACGATTTGCGAAGTCATGCGCTGCTCTTCTAACCTGGATCACAAG 912
 Db 224 GlnGluGlyValValLeuLysIleAsnGlyThrValThrGluAsnMetSerLeuThr 243
 Qy 913 GAGGCTGTAATGTCTGTAATAATAGCGCAGCGCTGACCATCTCCATTTGTAGCTGAGCT 972
 Db 244 AspAlaLysThrLeuIleGluArgSerLysGlyLysLeuLysMetValValGlnArgAsp 263
 Qy 973 GGCGGGAGCTGTTTCATG----- 990
 Db 264 GluArgAlaThrLeuLeuAsnValProAspLeuSerAspSerIleHisSerAlaAsnAla 283
 Qy 991 ACAGACCGGGAGCGGTGGCGAGGCGCGCGAG-----CGTGAG 1029
 Db 284 SerGluArgAspAspIleSerGlnSerLeuAlaSerAspHisSerGlyArgSer 303
 Qy 1030 CTGCGAGCGGAGGAGCTTCTCATGCGAGCGGTG----- 1065
 Db 304 HisAspArgProProArgArgSerGlnSerAspProAspGlnArgSerGluProSer 323
 Qy 1066 -----GCGATGGAGTCCACAGATCTCCAGGACGACGAGGAG 1104
 Db 324 AspHisSerThrGlnSerProGlnGlnProSerAsnGlySerLeuArgSerArgGluGlu 343
 Qy 1105 ATGAGCGCGCAAGGAGAAAGAAATGCCCCAGCAGCAGGAAATGAGAGATAC 1164
 Db 344 GluArgMetSerLysProGlyAlaIleSerThrProValLysHisValAspAspHisPro 363
 Qy 1165 CGGAAGGAGATGGAACAGATT---GTAGAGGAGGAGAGAGATTAAAGAACCAATGGAA 1221
 Db 364 ProLysAlaValGluGluValThrValGluLysAsnGluLysGlnThrProThrLeuPro 383
 Qy 1222 GAAGACTGGGCTCAAGAGAACAGCTACTCTGCTAAAAACCATCATCTGCTGAGGTACAC 1281
 Db 384 Glu-----ProLysProValTyrAlaGlnValGly 393
 Qy 1282 -----CCAGTACCCCTTCGCAAGCCCAAGTCAATTCGCGAAATATGA 1323
 Db 394 GlnProAspValAspLeuProVal-----SerPro-SerAsp----- 405
 Qy 1324 GGAAGGCTTTGACCCCTACTCTATGTTCCACCCAGACGACATCATGGGAAGGATGTCG 1383
 Db 406 ---GlyAlaLeuProAsnSerAlaHis---GluAspGlyIleLeuArgProSerMetLys 423
 Qy 1384 GCTCTACGATCAAGAGGAGGAGTCTTAGACCTGGCCCTGGAAGCGGTGTGGATC 1443
 Db 423 sLeuValLysPheArgLysGlyAspSerValGlyLeuArgLeuAlaGlyGlyAsnAsp-- 442
 Qy 1444 CCCCATTTGGGAAGGTGTGCTTCTGCTGTATGACGGGGGAGCTGCTGAGCGGCGATGG 1503

```

[4]
RN SEQUENCE OF 81-401 FROM N.A.
RP TSSUP=Brain;
RL Brennan J.E., Bredt D.S., Parkinson J.F., Manzana W.P., McClary J.A.;
RA submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
CC -! SUBUNIT: Interacts with DLGAP1/GKAP (By similarity). Is part of a
CC complex with DGAP1/GKAP, SHANK1 SHANK2 or SHANK3 (By similarity).
CC -! SUBCELLULAR LOCATION: Cytoplasmic. Concentrated at synaptic
CC junctions primarily on the presynaptic side. Also found in
CC postsynaptic density of neuronal cells (By similarity).
CC -! TISSUE SPECIFICITY: BRAIN.
CC -! SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -! SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -! SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -! SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; U83192; AAC52113.1; -.
EMBL; AF156495; AAD56173.1; -.
EMBL; U68138; AAB07736.1; -.
HSPG; P31016; IBE9.
Genev; HGNC:2903; DLG4.
MIM; 602887; -.
DR InterPro; IPRO00619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM0072; Gukc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 108 .. 194 PDZ 1.
FT DOMAIN 203 .. 289 PDZ 2.
FT DOMAIN 356 .. 436 PDZ 3.
FT DOMAIN 471 .. 541 SH3.
FT DOMAIN 577 .. 767 GUANYLATE_KINASE.
FT CONFLICT 46 .. 46 E -> V (IN REF. 3.).
FT CONFLICT 81 .. 83 VIV -> EFR (IN REF. 4.).
FT CONFLICT 399 .. 401 GDQ -> AGI (IN REF. 4.).
SQ SEQUENCE 767 AA; 85429 MW; BEI019159E85B2D8 CRC64;

Alignment Scores:
Pred. No.:      2 85e-06      Length:      767
Score:          234.50       Matches:     144
Percent Similarity: 38.10%    Conservative: 104
Best Local Similarity: 22.12% Mismatches:   210
Query Match:     5.88%      Indels:      193
DB:              1         Gaps:        34

US-09-502-945-5 (1-2162) x DLG4_HUMAN (1-767)
Qy 340 TCACAGGAAGTGAAAGGATGGCTGTGCACGCCCGACGGCCTCGCGCTGACT 399
Db 198 AlaclutysValMetGlulleLysLeulleLys---GlyProLysglyLeuGlyPheSer 216
Qy 400 GTGCGTGGTGGC-----CTGAGAGTTTGGCTGTGGGCTTCATCTCCACC 444

```

Db 217 IleAlaGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 236
 QY 445 CFCATCAAGGGGTCAGCAGACAGCTCGG---CTCCAGGTAGGGAGCAGATCGTC 501
 Db 237 IleIleGluGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 256
 QY 502 CGGATCAATGATATTCATCTCCCTCTACCCATGAGGAGTCAATCAACCTATTGGA 561
 Db 257 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaLeuLys 276
 QY 562 ---ACCAAGAAACGTGTCCATCAAGTG-----AGACATCGGCTGTATCCCC 609
 Db 277 AsnThrTyrAspValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 296
 QY 610 GTGAAAGCTCTCCGTAGCAGCCCTCACTGG---CAGTATGTGATCAGTTGTGTCG 666
 Db 297 SerTyrAlaProProAspIleThrThrSerTyrSerGlnHisLeuAspAsnGluLys 316
 QY 667 GAATCTGGGGGCTCGAGGACCTGGGC----- 696
 Db 317 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 332
 QY 697 ---TCCCTCGAATCGGAA-----AACAG 720
 Db 333 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluAspIleProArg 352
 QY 721 GAGAAGAGGTCTTCATCAGCTGTAGCTCCGAGGCTTGGCTGACGATTTCCAGC 780
 Db 353 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 372
 QY 781 GGCCCCATCCAGAGCTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
 Db 373 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 391
 QY 832 CTCTCTGCTGAGGTGGATGAGATGAGGGACCATGTTCTGAGTCAATGGCTGCAG 891
 Db 392 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 409
 QY 892 TTCTCTCACTGGATCACAAGAGGCTGTAATGTCTGAAATAAGCCGACCTGACC 951
 Db 410 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 428
 QY 952 ATCTCCTATGTAGTCACTGGCGGAGCTGTTTCAACAGACAGCGGAGCGGTGGCA 1011
 Db 429 ValThrIleIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 443
 QY 1012 GAGCGCGGACGCTGAGCTGAGCGGACGAGCTTCTCATCAGAGCGGCTGGCGATG 1071
 Db 444 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 461
 QY 1072 GAGTCCACAAGATCTCCAGGACGACGAGGAG-----ATGGAG 1110
 Db 462 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 481
 QY 1111 CGCAAGAGGAGAAAGAAATGCCCAGAGGCGAGAGAGAAATGAGAGATACCGGAAG 1170
 Db 482 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 501
 QY 1171 GAGATGGAACAGATCTAGAGGAGGAGAGAGAACTTT----- 1206
 Db 502 ValLeuHisValIleAspAlaSerAspGluGluTyrTrpGlnAlaArgValHisSer 521
 QY 1207 -----AAGAAGCAA 1215
 Db 522 AspSerGluThrAspAspIleGlyPheIleProSerLysArgValGluArgArgGlu 541
 QY 1216 TGGGAA-----GAAGACTGGGGCTCA-----AAGGAA 1242
 Db 542 TrpSerArgLeuLysAlaLysAspTrpGlySerSerSerGlySerGlnGlyArgGluAsp 561
 QY 1243 CAGCTACTCTTGCTAAACCATCACT---GCTGAGGTACACCCAGTACCCCTCCCAAG 1299
 Db 562 SerValLeuSerTyrGluThrValThrGlnMetGluValHisTyrAlaArg----- 578

QY 1300 CCAAAGTCATTTCCGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTTCACCCCA 1359
 Db 579 -----ProIleIleIleLeuLeuProThr 586
 QY 1360 GCAGATCATGGGGAAGATGTCCGGCTCCACGATCAAGAGGAGGATCCITAGACCT 1419
 Db 587 LysAspArgAla-----AsnAspAspLeuLeuSerGlu 597
 QY 1420 GGCCCTGGAAGCGGTGTGGACTCCCCCATTTGGGAAGTGGTCTTCTGCTGTGTATGA 1479
 Db 598 PheProAspLys----- 601
 QY 1480 GCGGGAGCTCTGAGCGGATGTGTCATTTGTAAAGG---GGACGAGATCATGGCAAT 1536
 Db 602 PheGlySerCysValPro-----HisThrThrArgProLysArgGluTyr-GluIle 618
 QY 1537 CAACGGCAAGATTGTGACAGACTACACCTGGCTGAGGCTGACGCTGCCCTGCAAGAGC 1596
 Db 618 eAspGlyArg-----AspTyrHisPheValSerSerArgGluLysMetGluLys-- 634
 QY 1597 CTGGAATCAGCGGGGAGTGGATCGACCTGTGTTGGCTCTGCCCTCCCAAGAGGATA 1656
 Db 634 ----- 634
 QY 1657 TGACGATGAGTACCTTCTTGTGAAGTCCAAAGGGGAAACCAATTCACGGCTTAGG 1716
 Db 635 -AspIleGlnAlaHisLysPheIleGluAlaGlyGlnTyrAsnSer---HisLeuTyrGlu 653
 QY 1717 AACAGT-----GAGCTCGGGCCCACTCTGCGTGAACAAA 1752
 Db 653 yThrSerValGlnSerValArgGluValAlaGluGlnGlyLysHisCysIleLeuAspVal 673
 QY 1753 GCCTCGGACCACTGTGAGAGGCCACATGACACACACAGATGGCATCTCTGGGACCT 1812
 Db 673 LserAlaAsnAlaValArgArgLeuGlnAlaAlaHisLeu-----HisProIleAlaIle 691
 QY 1813 GAATCTATCACCAGGAATCTCAAACTCCCTTGTGGCCCTGAAC----- 1855
 Db 691 ePheIleArgProArgSerLeuGluAsnValLeuGluIleAsnLysArgIleThrGluGlu 711
 QY 1856 -CAGGGCCAGATAGGAACACAGCTCGGGCCACTTTTTCAG- 1895
 Db 711 uGln-AlaArgLysAlaPheAspArgAlaThrLysLeuGluGlnGluPheThrGluCysP 731
 QY 1896 -----GCCAATGTGGAGGAAAGGAGCAGCGCTTTGGGAGAGATCAAGG---A 1947
 Db 731 hEserAlaIleValGluGlyAspSerPheGluGluIleTyrHisLysValLysArgValI 751
 QY 1948 TCCAGACTCTCATTT---CCTTCTCTCTGG 1973
 Db 751 leGluAspLeuSerGlyProTyrIleTrp 760
 RESULT 9
 DLG4_RAT STANDARD; PRT; 724 AA.
 AC P31016; p97631;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)
 DE (Synapse-associated protein 90) (Discs, large homolog 4).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=93040233; PubMed=1419001;
 RA Cho K.-O., Hunt C.A., Kennedy M.B.;
 RT "The rat brain postsynaptic density fraction contains a homolog of

[illegible]

```
QY 952 ATCTCCATTGTAGTCAGCTGGCGGGAGCTGTTTCATGACAGACGCGGAGCGGCTGSCA 1011
Db 386 ValThrIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
QY 1012 GAGGCGCGCAGCTGAGCTGCGGCGGAGGAGCTTCTCATGACAGAAAGCGGCTGCGCATG 1071
Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
QY 1072 GAGTCCACACAGATCTCCAGGACGACGAGGAG-----ATGGAG 1110
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
QY 1111 CGCGAAAGGAGAGAAAGAAATTCGCCAGAGGAGGAGGAGGAGGAGATACCGGAG 1170
Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 458
QY 1171 GAGATGGAACAGATGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1206
Db 459 ValLeuHisValIleAspAlaGlyAspGluGluTyrTrpGlnAlaArgValHisSer 478
QY 1207 -----AAGAGCA 1215
Db 479 AspSerGluThrAspAspIleGlyPheIleProSerLysArgValGluArgGlu 498
QY 1216 TGGGAA-----CAAGACTGGGCTCA-----AAGGAA 1242
Db 499 TrpSerArgLeuLysAlaLysAspTrpGlySerSerGlySerGlnGlyArgGluAsp 518
QY 1243 CAGCTACTCTGCTTAAACCATCACT-----GCTGAGGTACACCCAGTACCCCTTCGCAG 1299
Db 519 SerValLeuSerTyrGluThrValThrGlnMetGluValHisTyrAlaArg-----535
QY 1300 CCAAGTGATTTCCGGAATATGAGGAGGCTTGGACCTACTCTATCTTACCCAGCAGA 1359
Db 536 -----ProIleIleLeuGlyProThr 543
QY 1360 GCAGATCATGGGAGGAGTTCGGCTCTAGCATCAAGAGGAGGAGGATCTTAGACCT 1419
Db 544 LysAspArgAla-----AsnAspAspLeuLeuSerGlu 554
QY 1420 GGCCTTGAAGGGGTGGAGTCCCCCATTTGGAGGTGGTCTGCTGTGTATGA 1479
Db 555 PheProAspLys-----558
QY 1480 GCGGGAGCTGCTGACGGCATGGTGCATTGTGAAGG---GGACGAGATCATGCAAT 1536
Db 559 PheGlySerCysValPro-----HisThrArgProLysArgGluTyr-GluI 575
QY 1537 CAACGGCAGATTTGACAGACTACACCTGCTGAGGTGAGCTGCCCTGCGAAGGC 1596
Db 575 eAspGlyArg-----AspTyrHisPheValSerSerArgGluLysMetGluLys-- 591
QY 1597 CTGGAATCAGGGGGGAGTGGATCGACCTTGTGGTGTGGCTGCTGCCCCCAAGAGTA 1656
Db 591 -----591
QY 1657 TGACGATGAGCTGACCTTCTTGTGAGTCCAAAGGGGAGAAACAAATTCACCGCTTAGG 1716
Db 592 AspIleGlnAlaHisLysPheIleGluAlaGlyGlnTyrAsnSer---HisLeuTyrGI 610
QY 1717 AAACAGT-----GAGCTCCGGCCCGACCTCGTGAACACAA 1752
Db 610 yThrSerValGlnSerValArgGluValAlaGluGlnGlyLysHisCysIleLeuAspVa 630
QY 1753 GCCTCGGACCGCTTGTGAGAGGCCACATGACACACAGATGGCATCCTTGGGACCT 1812
Db 630 lSerAlaAsnAlaValArgLeuGlnAlaAlaHisLeu-----HisProIleAlaI 648
QY 1813 GATCTATCACCGAGTAATCTCAACTCCCTTGGCCCTCGAAC-----1855
Db 648 ePheIleArgProArgSerLeuGluAsnValLeuGluIleAsnLysArgIleThrGluI 668
```

```
QY 1856 -CAGGGCCAGATAGGAACAGCTCGGGCCACTTTTGAAG-----1895
Db 668 uGln-AlaAgLysAlaPheAspArgAlaThrLysLeuGluGlnGluPheThrGluCysp 688
QY 1896 -----GCCAATGTGGAGAAAGGAGGAGCAGCGCGTTTGGAGAGATCTCAAGG---A 1947
Db 688 heSerAlaIleValGluGlyAspSerPheGluGluIleTyrHisLysValLysArgValI 708
QY 1948 TCCAGACTCTCATTT---CCTTCTCTCTGG 1973
Db 708 leGluAspLeuSerGlyProTyrIleTrp 717

RESULT 10
DLG3_HUMAN
ID DLG3_HUMAN STANDARD; PRT; 817 AA.
AC Q92796; O9ULI8;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic protein SAPI02 (Synapse-associated protein 102)
DE (Neuroendocrine-DLG) (NE-DLG) (Discs, large homolog 3).
GN DLG3 OR KIAA1232.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=97332623; PubMed=918857;
RA Makino K., Kuwahara H., Masuko N., Nishiyama Y., Morisaki T.,
RA Sasaki J., Nakao M., Kuwano A., Nakata M., Ushio Y., Saya H.;
RT "Cloning and characterization of NE-dlg: a novel human homolog of the
RT Drosophila discs large (dlg) tumor suppressor protein interacts with
RT the APC protein.";
RL Oncogene 14:2425-2433(1997).
RN [2]
RP SEQUENCE OF 330-817 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
CC - FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC - SUBUNIT NR2B (BY SIMILARITY).
CC - SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC - SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC - SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U49089; AAB61453.1; -.
CC EMBL; AB033058; BAA86546.1; -.
CC HSSP; Q12959; 1PDR.
CC Genew; HGNC:2902; DLG3.
CC MIM; 300189; -.
CC InterPro; IPR00619; Guanylate_kin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00595; PDZ; 3.
CC Pfam; PF00625; Guanylate_kin; 1.
CC ProDom; PD000066; SH3; 1.
```

[illegible]

Db 559 AspLeuArgGluGln.....MetMetAsnSerSerIleSerGlySerGlySerLeu 576

Qy 1087 CTCCAGGAGCAGCAG 1101

Db 577 ArgThrSerGlnLys 581

RESULT 12

DLG4_MOUSE	STANDARD;	PRT;	724 AA.
ID	DLG4_MOUSE		
AC	Q62108;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last annotation update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)		
DE	(Synapse-associated protein 90 (PSD-95) (Discs, large homolog 4).		
GN	DLG4 OR DLGH4 OR PSD95.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=DBA/2; TISSUE=Brain;		
RA	Kohmura N., Yagi T.;		
RL	Submitted (MAY-1995); to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR		
CC	SUBUNITS. MAY BE INVOLVED IN SYNAPTONEGENESIS.		
CC	-1- SUBUNIT: Interacts with DLGAP1/GKAP (By similarity). Is part of a		
CC	complex with DLGAP1/GKAP, SHANK1 SHANK2 OR SHANK3 (By similarity).		
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic. Concentrated at synaptic		
CC	junctions primarily on the presynaptic side. Also found in		
CC	postsynaptic density of neuronal cells (By similarity).		
CC	-1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.		
CC	-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN		
CC	-1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.		
CC	-1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcement/		
CC	or send an email to license@isb-sib.ch).		
EMBL;	D50621; BAA09297.1;		
DR	HSSP; P31016; LBE9		
DR	MGD; MGI:1277959; Dlg4.		
DR	InterPro; IPR000619; Guanylate_kin.		
DR	InterPro; IPR001478; PDZ.		
DR	InterPro; IPR001452; SH3.		
DR	Pfam; PF000018; SH3; 1.		
DR	Pfam; PF000595; PDZ; 3.		
DR	Pfam; PF00625; Guanylate_kin; 1.		
DR	ProDom; PD000066; SH3; 1.		
DR	SMART; SM00072; GUKC; 1.		
DR	SMART; SM00228; PDZ; 3.		
DR	SMART; SM00326; SH3; 1.		
DR	PROSITE; PS00856; GUANYLATE_KINASE_1; 1.		
DR	PROSITE; PS50052; GUANYLATE_KINASE_2; 1.		
DR	PROSITE; PS50106; PDZ; 3.		
DR	PROSITE; PS50002; SH3; 1.		
KW	SH3 domain; Repeat.		
FT	DOMAIN 65 151 PDZ 1.		
FT	DOMAIN 160 246 PDZ 2.		
FT	DOMAIN 313 393 PDZ 3.		
FT	DOMAIN 428 498 SH3.		
FT	DOMAIN 534 724		
SQ	SEQUENCE 724 AA 80472 MW; 7EFCC99E1FF90BA CRC64;		
	GUANYLATE KINASE.		

Alignment Scores:

Pred. No.:	5.46e-06	Length:	724
Score:	229.50	Matches:	146

Percent Similarity:	37.54%	Conservative:	101
Best Local Similarity:	22.19%	Mismatches:	204
Query Match:	5.75%	Indels:	207
DB:	1	Gaps:	35

US-09-502-945-5 (1-2162) x DLG4_MOUSE (1-724)

[illegible]

```
Db 459 ValLeuHisValIleAspAlaSerAspGluGluTrpTrpGlnAlaArgValHisSer 478
QY 1135 CAGAAGGCAGCAGCAAAATGAGAGATACCGGAAGGAGATGGAACACAGATGTAGAGGAG 1194
Db 479 AspSerGluThrAsp-----AspIleGlyPheIleProSerLys 491
QY 1195 GAAGAGAAGTTTAAGAAGCAATGGAA-----GAAGACTGGGGCTCA----- 1236
Db 492 ArgArgValGluArgArgGluTrpSerArgLeuLysAlaLysAspTrpGlySerSer 511
QY 1237 -----AAGAACAGACTACTCTGTCCTAAACCATCACT---GCTGAGGTA 1278
Db 512 GlySerGlnGlyArgGluAspSerValLeuSerTyGluThrValThrGlnMetGluVal 531
QY 1279 CACCAGTACCCCTTCGCAAGCAAAAGTATTCCGGAAATATGAGGAAGGCTTTGACC 1338
Db 532 HisTyAlaArg-----Pro 536
QY 1339 CTACTCTATGTTTACCCAGCAGATCATGCGGAAGGATGTCGGCTCCTACGCATCAA 1398
Db 537 IleIleIleLeuGlyProThrLysAspArgAla----- 547
QY 1399 GAAGAGGAGTCTTAGACCTGGCTGGAAGCGGTGGAGTCCCCCATTTGGAAGGT 1458
Db 548 AsnAspAspLeuLeuSerGluPheProAspLys----- 558
QY 1459 GCTCGTTTCTGCTGTATGATGCGGGGAGCTGCTGAGCGGCATGTGGCATTTGAAAGG 1518
Db 559 -----PheGlySerCysValPro-----HisThrThrArg 568
QY 1519 ---GGACGAGATCATGGCAATCAACGCGCAAGATTGTGACAGACTACACCTGGCTGAGGC 1575
Db 569 ProLysArgGluTyr-GluIleAspGlyArg-----AspTyrHisPheValSerSe 585
QY 1576 TGACGCTGCCCTGCAGAAAGCCCTGGAAATCAGCGGGGAGTGGATGAGTGGCTGTGCTGC 1635
Db 585 rArgGluLysMetGluLys----- 591
QY 1636 CGTCTCCCCCAAGGAGTATGACGATGACGTGACCTTCTGCTGAAGTCCAAAGGGG 1695
Db 592 -----AspIleGlnAlaHisLysPheIleGluAlaGlyGlnTy 604
QY 1696 AACCAAAATTCACGCGTTAGGAACAGT-----GAGCTCCG. 1731
Db 604 rAsnSer---HisLeuTyrGlyThrSerValGlnSerValArgGluValAlaGluGlnG 623
QY 1732 GCCCCACCTCGTGAACAAAGCCCTCGACCCCTTGAGAGGCCACATGACACACAC 1791
Db 623 yLysHisCysIleLeuAspValSerAlaAsnAlaValArgLeuGlnAlaAlaHisLe 643
QY 1792 CAGATGGCATCTTGGGACCTGAATCTATCACCAGGAATCTCAAACTCCCTTTGGCCCT 1851
Db 643 u-----HisProIleAlaIlePheIleArgProArgSerLeuGluAsnValLeuGlu 661
QY 1852 GAAC-----CAGGCCAGATAAGAAACAGCTCGGGCCACTTTTTCGA 1893
Db 661 eAsnLysArgIleThrGluGluGln-AlaArgLysAlaPheAspArgAlaThrLysLeuG 681
QY 1894 AG-----GCCAATGTGAGGAAGGAGGAGCCAGCCGCTTT 1929
Db 681 luGlnGluPheThrGluCysPheSerAlaIleValGluGlyAspSerPheGluGluIle 701
QY 1930 GGGAGAAGATCTCAAGG---ATCCAGACTCTCATT---CCTTTCCCTCTGG 1973
Db 701 yrHisLysValLysArgValIleGluAspLeuSerGlyProTyrIleTrp 717
```

RESULT 13

DLGL_RAT	STANDARD;	PRT;	911 AA.
ID	DLGL_RAT		
AC	O62656;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		

Alignment Scores:
Pred. No.: 9,04e-06 Length: 911
Score: 226.00 Matches: 120
Percent Similarity: 37.32% Conservative: 83
Best Local Similarity: 22.06% Mismatches: 180
Query Match: 5.67% Indels: 161
DB: 1 Gaps: 24

US-09-502-945-5 (1-2162) x DLGI_RAT (1-911)

QY	343	AGGAAGCTGAAGGAGGTGCGTCTGACCGCTGCACCCGAAAGCGCTCGCCCTGACTGTG	402
Db			
Db	314	ArgLysAsnHisGluIleLysLeuIleLys---	332
QY	403	CTGGTGGC-----CTGAGATTGGCTGTGGCTCTTCATCTCCACACCTC	447
Db			
Db	333	AlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLysIle	352
QY	448	ATCAAAAGCGCTCAGGCAGACAGCTCGG---CTCCAGTAGGGACAGATCTGTCGG	504
Db			
Db	333	IleGluGlyAlaAlaHisLysAspGlyLysGlnIleGlyAspLysLeuAla	372
QY	505	ATCAATGATATTCATCTCTCTCTGTGACCATGAGAGGTCAACCTCATTCGA---	561
Db			
Db	373	ValAsnSerValCysLeuGluValThrHisGluGluAlaValThrAlaLeuLysAsn	392
QY	562	ACCAAGAAACTGTCTCCATCAAGTCAGA-----CACATCGGC	600
Db			
Db	393	ThrSerAspPheValTyrLeuLysAlaAlaLysProThrSerMetTyrIleAsnAspGly	412
QY	601	CTGATCCCGCTGAAAAGCTCTCCTGAT---GAGCCCTCCTCAGTGGCAGATCTGATCATG	657
Db			
Db	413	TyrAlaPro-----ProAspIleThrAsnSerSerSerGlnSerValAspAsn	428
QY	658	TTGTGTGCGNATCTGGGGCGTCGAGCAGCGCTGGGCTCCCT---	702
Db			
Db	429	HisValSerProSerTyrLeuGlyGlnThrProAlaSerProAlaArgTyrSerPro	448
QY	703	-----GGAATCTGGGAA---AACAAAGGAGAAGGTCTTCATCAGC	741
Db			
Db	449	IleSerLysAlaValLeuGlyAspGluThrArgGluProArgLysValValLeu	468
QY	742	CTGTAGCTCCCGAGCGCTTGGCTGCAGCATTTCCAGCGGGCCCCATCCAGAAGCCTCGC	801
Db			
Db	469	HisArgGlySerThrGlyLeuGlyPheAsnIleValcylGlyGlu---AspGlyGluGly	487
QY	802	ATCTTTATCGCCATNG-----AAACCTGGCTCCTGCTGCTGAGGTGGGATG	852
Db			
Db	488	IlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu-----Leu	505
QY	853	GAGATAGGGACCATGTCGAAAGTCAATGGCTGCAGCTTCTCTAACTGGATCACAAAG	912
Db			
Db	506	ArgLysGlyAspArgIleIleSerValAsnSerValAspLeuArgAlaAlaSerHisGlu	525
QY	913	GAGCTGTAATGTGTGAAAAATACCGCAGCGCTGACCATCTCCATTGTAGTCGACGT	972
Db			
Db	526	GlnAlaAlaAlaLeuLysAsnAla---GlyGlnAlaValThrIleValAlaGlnTyr	544
QY	973	GGCGGGAGCTGTTTCATGACAGACCGGGAGCGGCTGGCAGCGCGCGCGAG-----	1023
Db			
Db	545	ArgProGluGluTyrSerArgPheGluAlaLysIleHisAspLeuArgGluThrMetMet	564
QY	1024	-----CGTAGCTCAGCGCAGGAGCTTCTC	1050
Db			
Db	565	AsnSerValSerSerGlySerLeuArgThrSerGlnLysArgSerLeuTyr	584
QY	1051	ATGCAGAAGCGGTGGCGATGGATCCACAG-----	1083
Db			
Db	585	ValArgAlaLeuPheAspTyrAspLysThrLysAspSerGlyLeuProSerGlnGlyLeu	604
QY	1083	-----	1083
Db			
Db	605	AsnPheLysPheGlyAspIleLeuHisValIleAsnAlaSerAspAspGluTyrTrpGln	624
QY	1084	-----ATCCTCCAGGACGACGAC	1101
Db			
Db	625	AlaArgGlnValThrProAspGlyGluSerAspGluValGlyValIleProSerLysArg	644
QY	1102	GAGATGGCGGCAAGGAGAAAGAAATGTCAGAGAGCGCAGAGGAAAAATGAGAGA	1161

Db	645	ArgValGluLysLysGluArgAlaArgLeu-----LysThrValLysPheAsnSerLys	662
QY	1162	TACCGAAGGAGATGGAAACAGATTGTAGAGGAGGAAGACAAGT-----TT	1206
Db	663	ThrArgGlyAspLysGlyGluIleProAspAsp-MetGlySerLysGlyLeuLysHisVa	682
QY	1207	AAGAAGCAATGGGAAGAACTGGGCTCAAGGAACACACTACTCTGCTCAAAACCATC	1266
Db	682	IThrSerAsnAlaSerAspSerGluSerSerTyHisGluTyGlyCysSerLysGlyG1	702
QY	1267	ACTGCTGAGGTACACC---CAGTACCCCTTCGCAACCAA-----	1303
Db	702	YcInGluGluTyrValLeuSerTyGluProValAsnGlnGluValAsnTyThrAr	722
QY	1304	-----AGTGATT	1311
Db	722	gProValIleLeuGlyProMetLysAspArgValAsnAspAspLeuLysSerGluPh	742
QY	1312	C-----CGAAATATGAGGA	1326
Db	742	eProAspLysPheGlySerCysValProHisThrThrArgProLysArgAspTyrgluVa	762
QY	1327	AGGCTTTGACCCCTACTCTATGTCACCCAGACAGATCATGGGAAGGATGTCGGCT	1386
Db	762	IaspGlyArgAspTyHisPheValThrSerArgGluGlnMetGluLysAspIleGlnG1	782
QY	1387	CCTACGCATCAAGAAGGAGGATCCTTAGACCTGGCCCTGGAAAGC---GGTGGGCACT	1443
Db	782	uHisLysPheIleGluAlaGlyGlnTyrsnAsnHisLeuTyrglyThrServalGlnSe	802
QY	1444	CCCATTTGGGAAGGTGCTGCTTCTGTGTATGAGCGGGAGCTGCTGACGGCATGG	1503
Db	802	r-----ValArgAlaValAlaGluLysGly-----LysHisCy	813
QY	1504	TGGCATTGTGAAGGGACGAGATCATGCATCAACGCCAAGATGTGCACAGACTACAC	1563
Db	813	s-----IleLeuAspValSerGlyAsnAlaIleLysArgLeuG1	826
QY	1564	CCTGGCTGAG	1573
Db	826	nIleAlaGln	829
RESULT	14		
ZOL_HUMAN			
ID	ZOL_HUMAN	STANDARD;	PRT; 1736 AA.
AC	Q07157;		
DT	01-OCT-1994 (Rel. 30, Created)		
DT	01-OCT-1994 (Rel. 30, Last annotation update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Tight junction protein ZO-1 (zonula occludens 1 protein) (zona		
DE	occludens 1 protein) (Tight junction protein 1).		
GN	TJPI OR ZO1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RA	MEDLINE=93361541; PubMed=8395056;		
RA	Willott E., Balda M.S., Fanning A.S., Jameson B., van Itallie C.,		
RA	Anderson J.M.;		
RT	"The tight junction protein ZO-1 is homologous to the Drosophila		
RT	discs-large tumor suppressor protein of septate junctions.;"		
RL	Proc. Natl. Acad. Sci. U.S.A. 90:7834-7838(1993).		
CC	-!- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL		
CC	REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY		
CC	HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN		
CC	MIGHT BE INVOLVED IN STABILIZING JUNCTIONS.		
CC	-!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-3.		
CC	-!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.		
CC	MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT		
CC	OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT		


```

QY 1373 -----AAGATGTCGGCTCTACGATCAAGAAGGAGG 1406
Db 399 InLeuMetLysMetGlyPheLeuArgProSerMetLysLeuValLysPheArgLysGlyA 419
QY 1407 GATCCTTAGACCTGGCCCTGGAGCGGTGTGGACTCCCATTTGGGAAGGTGGTCTTT 1466
Db 419 sPserValGlyLeuArgLeuAlaGlyGlyAsnAsp-----ValGly---IlePheValA 436
QY 1467 CTGCTGTATGAGCGGGAGCTGTGAGCGGCATGGTGGCATTGTGAAGGGGACGAGA 1526
Db 436 laGlyValLeuGluAspSerProAlaIaLysGlu---GlyLeuGluGluGlyAspGlnI 455
QY 1527 TCATGCCAATCAAGCGCAAGATGTGACAGACTACACCTGGCTGAGCGTGCAGCTGCC 1586
Db 455 leLeuArgValAsnAsnValAspPheThrAsnIleIleArgGluGluAlaValLeuPheL 475
QY 1587 TCGAGAAGCCCTGGAATCATGAGCGGGAGCTGATCGACCTTGTGGTCCGCTCTGCCCC 1646
Db 475 eu-----LeuAspLeu-----p 479
QY 1647 CAAGGACTATGACGATGAGCTGACCTTCTGCTGAAGTCCAAAGGGGAACCAAAATTC 1706
Db 479 roLys-----GlyGluGluValThrIleLeuAlaGlnLysLysLys-----492
QY 1707 AGCGCTTAGAACAAGTGTGCGGCCGCCACCTCTGCTGAACACACAAGCTCGGACGACC 1766
Db 493 -----AspValTyArgArgIleValGluSerAspValGlyAspSerp 507
QY 1767 TTGAG---AGAGGCCACATGACACACAGATGCATCCTTGGGACCTGATCTATCATC 1823
Db 507 heTyIleArgThrHisPheGluTyArgLysGluSerProTyrgly-----522
QY 1824 CCAGGAATCTCAACCTCTTGGCCCTGAACCGAGGCGCAGATAAGGAACAGCTCGGGCC 1883
Db 523 -----LeuSerPheAsnLysGlyGluValPheArgIaValAsp 536
QY 1884 ACTTTTGAAGGCAATGTGGAGGAAGGAGGAGCAGCGCTTTGGGAGAAGATCTCA 1943
Db 536 hrLeuTyArgAsnGlyLysLeu---GlySerTrpLeuAlaIleArgIleGlyLysAsnHisL 555
QY 1944 AGGATCCAGACTCTCATCTTCTCT-----CTGGCCCAAGTGAA 1982
Db 555 ysgluValGluArgGlyIleleProAsnLysAsnArgAlaGluGlnLeuAlaSerValG 575
QY 1983 TTTGCTCTCTCCCA-----GCTTTGGGGGAC 2008
Db 575 InTyThrLeuProLysThrAlaGlyGlyAsp 585

RESULT 15
ID DLG2_RAT STANDARD; PRT; 852 AA.
AC Q63622; Q62939; P70548;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Channel associated protein of synapse-110 (Chapsyn-110) (Synaptic
DE density protein PSD-93) (Discs, large homolog 2).
GN DLG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96310881; PubMed=8755482;
RA Kim E., Cho K.O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of
RL Chapsyn-110, a member of the PSD-95 family of proteins.";
RN Neuron 17:103-113(1996).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96193770; PubMed=8625413;
RA Brenman J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E.,

```

```

RA Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F.,
RA Froehner S.C., Bredt D.S.;
RT "Interaction of nitric oxide synthase with the postsynaptic density
protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
RN Cell 84:757-767(1996).
[3]
RP SEQUENCE FROM N.A.
RA Irie M., Hata Y., Takai Y.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS AS WELL AS POTASSIUM CHANNELS.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U49049; AAB53243.1; -
CC EMBL; U50717; AAC52643.1; -
CC EMBL; U53368; AAB48562.1; -
CC HSPS; Q12959; IPDR.
CC InterPro; IPR000619; Guanylate_kin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00595; PDZ; 3.
CC Pfam; PF00625; Guanylate_kin; 1.
CC ProDom; PD000066; SH3; 1.
CC SMART; SM00072; GuK; 1.
CC SMART; SM00228; PDZ; 3.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
CC PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
CC PROSITE; PS0106; PDZ; 3.
CC PROSITE; PS00002; SH3; 1.
CC SH3 domain; Repeat.
KW DOMAIN 98 184 PDZ 1.
FT DOMAIN 193 279 PDZ 2.
FT DOMAIN 421 501 PDZ 3.
FT DOMAIN 536 606 SH3.
FT DOMAIN 662 852 GUANYLATE_KINASE.
FT CONFLICT 181 182 VR -> IL (IN REF. 2).
FT CONFLICT 228 228 I -> M (IN REF. 2).
FT CONFLICT 326 326 R -> K (IN REF. 2).
FT CONFLICT 339 339 D -> E (IN REF. 3).
FT CONFLICT 450 454 MISSING (IN REF. 2).
FT CONFLICT 464 465 GD -> RK (IN REF. 2).
FT CONFLICT 474 474 D -> H (IN REF. 2).
FT CONFLICT 476 476 R -> P (IN REF. 2).
FT CONFLICT 478 478 A -> D (IN REF. 2).
FT CONFLICT 484 486 AAA -> LP (IN REF. 2).
FT CONFLICT 506 506 A -> S (IN REF. 2).
FT CONFLICT 569 569 H -> N (IN REF. 2).
FT CONFLICT 586 586 L -> Q (IN REF. 2).
FT CONFLICT 626 641 MISSING (IN REF. 2).
FT CONFLICT 639 639 K -> A (IN REF. 3).
FT CONFLICT 726 726 F -> L (IN REF. 1).
FT CONFLICT 733 733 N -> Y (IN REF. 2).
FT CONFLICT 749 749 E -> V (IN REF. 1).
FT CONFLICT 756 756 L -> H (IN REF. 2).
FT CONFLICT 791 792 KR -> NG (IN REF. 2).
FT CONFLICT 794 794 T -> M (IN REF. 2).
SQ SEQUENCE 852 AA; 94934 MW; F8D414A8B9CF5B09 CRC64;

```

Alignment Scores:
Pred. No.: 3.58e-05 Length: 852

Score: 215.50 Matches: 119
Percent Similarity: 39.50% Conservative: 86
Best Local Similarity: 22.93% Mismatches: 188
Query Match: 5.40% Indels: 126
DB: 1 Gaps: 24

US-09-502-945-5 (1-2162) x DLG2_RAT (1-852)

```
QY 307 CAGGTGGGAATATGATCAGCTGACCCCGCGCTCCAGGAGCTCAAGGAGGTGGCTGTG 366
Db 91 GtulleGluTyrGlu.....PheGluGluIleThrLeu 101
QY 367 GACCGTCTGCACCCCAAGGCTCGGCTGAGTGTGCGTGGCTGGAG----- 417
Db 102 GluArgGlyAsn---SerGlyLeuGlyPheSerIleAlaGlyGlyThrAspAsnProHis 120
QY 418 -----TTTGGCTGTGGCTCTCATCTCCACCTCATCAAGGCGGTGACGACAGC 471
Db 121 IleGlyAspProGlyIlePheIleThrLysIlePheGlyGlyAlaAlaGlu 140
QY 472 GTCGGG---CTCCAGGTAGGGAGAGATCGTCGGGATCAATGGATATTCATCTCTCC 528
Db 141 AspGlyArgLeuArgValAsnAspCysIleLeuArgValAsnGluValAspValSerGlu 160
QY 529 TGTACCCATGAGAGGTCAACCTCATTCGA---ACCAAGAAACTGTGTCCATCAA 585
Db 161 ValSerHisSerLysAlaValGluAlaLeuLysGluAlaGlySerIleValArgLeuTyr 180
QY 586 GTCGACACATCGGCTGATCCCGTGAAGCTCTCCTGATGAGCCCTCCTACTGGCAG 645
Db 181 ValArgArg----- 183
QY 646 TATGTGGATCAGTTTGTGTGGAATCTGGGGGTGCGGAGCGCTGGCTCCCTGGA 705
Db 183 ----- 183
QY 706 AATCGGAAACAGAGAGAGAGTCTTCATCAGCCTGGTA---GGCTCCCGAGCCCTT 762
Db 184 ArgArgProIleLeuGluThrValValGluIleLeuLysGlyProLysGlyLeu 203
QY 763 GGCTGCGAGATTCACGGCGCCCTCCAGAG---CCTGGC-----ATCTTT 807
Db 204 GlyPheSerIleAlaGlyValGlyValGlyAsnGlyIleProGlyAspAsnSerIleTyr 223
QY 808 ATCAGCATGTGAACCTGCTCCCTGCTGCTGAGTGGGA---TTGGAGATAGGGGAC 864
Db 224 ValThrLysIleLeuAspGlyGlyAlaAlaGlnLysAspGlyArgLeuGlnValGlyAsp 243
QY 865 CAGATTGCTGAAGTCAATGGCGTGGCTCTCTAATGATGATCAGAGGAGGCTGTAAT 924
Db 244 ArgLeuLeuMetValAsnAsnTyrSerLeuGluGluValThrHisGluAlaValAla 263
QY 925 GTGCTGAATAATAGCCCGACCTGACATCTCCATCTGCTGAGTGGCGCGGAGCTG 984
Db 264 IleLeuLysAsnThr---SerAspValValTyrLysValGlyLysProThrThrIle 282
QY 985 TTCATGACAGACCGGCGGCTGGCAGAGCGCGGAGCGGTGAGTGCAGCGCGAGGAG 1044
Db 283 TyrMetThrAspProTyrGlyProProAspIle-----ThrHisSerTyrSer 298
QY 1045 CTTCTCATGACAGCGGCTGGGATGGATGATCCCAAGATCTCCAGGAGCAGCAGGAG 1104
Db 299 ProProMetGluAsnHisLeuLeuSerGlyAsnAsnGlyThrLeuGluTyrLysThrSer 318
QY 1105 ATG-----GAGCGGCAAGGAGAAAGAAATTCGCCAGAGGCGCAGAGGAAAT 1155
Db 319 LeuProIleSerProGlyArgTyrSerProIleProLysHisMetLeuValGluAsp 338
QY 1156 GAGAGATACCGGAGGAGATGGACATCTAGAGGAGGAGAGAGATTAAGAACAA 1215
Db 339 Asp---TyrThrArgProProGluProValTyrSerThrValAsnLys-LeuCysAsp-- 356
QY 1216 TGGGAAGAAGACTGGGCTCAAGGAGACAGCTACTCT-----TGCCTAAACCATC 1266
```

```
Db 357 ----LysProAlaSerProArgHis---TyrSerProValGluCysAspLys---Se 372
QY 1267 ACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATTCCGGAATATGAGGA 1326
Db 372 rPheLeuLeuSerThrProTyrPro-----HisTyrHisLe 384
QY 1327 AGGCTTTGACCCCTACTCTATGTTACACCCACAGAG----- 1360
Db 384 uGlyLeuLeuProAspSerAspMetThrSerHisSerGlnHisSerThrAlaThrArgG 404
QY 1361 -----CAGATCATGGGGAAGGATGTCGGCTCCCTACG 1392
Db 404 nProSerValThrLeuGlnArgAlaIleSerLeuLeuGlyGluProArgLysValValLe 424
QY 1393 CATCAAGGAAGGAGGATCCTTAGACCTGGCCCTGGAAGCGGTGTGGACTCCCCCATTTGG 1452
Db 424 uHisLysGlySerThrGlyLeuGlyPheAsnIleValGlyGlyGluAspGlyGlu----- 442
QY 1453 GAAGTGTCTCTTCTGCTGTGTATGACCGGGGAGCTGCTGAGCGGCATGGTGGCATTTGT 1512
Db 443 -GlyIlePheValSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGluGlu 462
QY 1513 GAAAGGGGACGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTGGCTGA 1572
Db 462 nArgGlyAspGlnIleLeuSerValAsnGlyIleAspLeuArgGlyAlaSerHisGluG 482
QY 1573 GGCTGACGCTGCTCCCTGACAGAGCGCTGGAATCAGGCGGGGAGTGGATCGACCTGTGGT 1632
Db 482 nAlaAlaAlaAlaLeu-----LysGlyAlaGlyGlnThrValThrIleI 497
QY 1633 TCCCTCTGCCCCCAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1692
Db 497 eAlaGlnTyrGlnProGluAspTyr-----AlaArgph 508
QY 1693 GGGAAACCAATTCACGGCTTAGGAAACAGTGCAGTCCGCGCCCACTCGTGAAC 1747
Db 508 eGluAlaLysIleHis-----AspLeuArgGluGlnMetMetAsn 521
```

Search completed: March 21, 2003, 12:48:39
Job time : 51.5761 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:20:42 ; Search time 74.1434 Seconds

(without alignments)
12016.549 Million cell updates/sec

Title: US-09-502-945-5

Perfect score: 3989

Sequence: 1 cctggccggtcggtgcg.....tatttccagcttaaaaaa 2162

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09502945/runat_14032003_101058_19113/app_query.fasta_1.10379
-DB=SPREMBL_21 -QFMT=fastan -MATRIX=spool -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -SUFFIX=rspt -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09502945 @CGN_1.1.565 @runat_14032003_101058_19113 -NCPU=3
-NO_XLPXY -NO_MMALP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_protist.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2558.5	64.1	548	11	Q9ES65	Q9es65 mus musculus

ID	Q9ES65	PRELIMINARY	PRT	548 AA
AC	Q9ES65;			
DT	01-MAR-2001 (TREMBlrel. 16, Created)			
DT	01-MAR-2001 (TREMBlrel. 16, Last sequence update)			
DE	01-DEC-2001 (TREMBlrel. 19, Last annotation update)			
DE	Harmonin isoform al.			
GN	2010016F01RIK OR USHIC.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20428180; PubMed=10973247;			
RA	Verpy E., Leblond M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,			
RA	Mansour A., Blanchard S., Kobayashi I., Keats B.J., Slim R., Petit C.,			
RT	"A defect in harmonin, a PDZ domain-containing protein expressed in			
RT	the inner ear sensory hair cells, underlies Usher syndrome type 1C.";			
RL	Nat. Genet. 26:51-55(2000).			
DR	EMBL; AF228924; AAG12457.1; -			
DR	HSSP; P29476; IOAV.			
DR	MGD; MGI:1919338; 2010016F01RIK.			
DR	InterPro; IPR001478; PDZ.			

ALIGNMENTS

RESULT 1

Q9ES65
ID Q9ES65
AC Q9ES65;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Harmonin isoform al.
GN 2010016F01RIK OR USHIC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20428180; PubMed=10973247;
RA Verpy E., Leblond M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,
RA Mansour A., Blanchard S., Kobayashi I., Keats B.J., Slim R., Petit C.,
RT "A defect in harmonin, a PDZ domain-containing protein expressed in
RT the inner ear sensory hair cells, underlies Usher syndrome type 1C.";
RL Nat. Genet. 26:51-55(2000).
DR EMBL; AF228924; AAG12457.1; -
DR HSSP; P29476; IOAV.
DR MGD; MGI:1919338; 2010016F01RIK.
DR InterPro; IPR001478; PDZ.

DR Pfam: PF00595; PDZ; 3.
 DR SMART: SM0228; PDZ; 3.
 DR PROSITE: PS0106; PDZ; 3.
 SQ SEQUENCE 548 AA; 61712 MW; 44A245AA32091070 CRC64;

Alignment Scores:

Pred. No.: 3,4e-177 Length: 548
 Score: 2558.50 Matches: 503
 Percent Similarity: 93.98% Conservative: 12
 Best Local Similarity: 91.79% Mismatches: 11
 Query Match: 64.14% Indels: 22
 DB: 11 Gaps: 1

US-09-502-945-5 (1-2162) x Q9ES65 (1-548)

QY 97 ATGACCAAAAGTGGCCCGAGATTCGGCATTAAGTGGATTTCTGATTGAATGAT 156
 |||||
 DB 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
 |||||
 QY 157 GCAGAGAAGGACTATCTCTATGATGTCTGCGAATGTACCAACAGACCATGGAGTGGCC 216
 |||||
 DB 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
 |||||
 QY 217 GTGCTCGTGGAGACCTGAAGCTGTGATCAATGAACACCGCTCTGCTCTGTTGAT 276
 |||||
 DB 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProAsnArgLeuProLeuPheAsp 60
 |||||
 QY 277 GCCATTCCGCGCTGATCCCACTGAAGCAGCAGGTTGGAATATGATCAGCTGACCCCGG 336
 |||||
 DB 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 |||||
 QY 337 CCTCCAGGAAGCTGAAGGAGTGGCTCTGACCGCTCTGACCCCGAAGGCTCGGCTG 396
 |||||
 DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
 |||||
 QY 397 AGTGTGGTGGTGGCTGGAGTTTGGCTGGCTGCTTCATCTCCCACTCATCAAGGC 456
 |||||
 DB 101 SerValArgGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
 |||||
 QY 457 GGTACGAGCAGAGCTGGCTCCAGTTCAGTGGGAGAGATCTCCGGATCAATGATAT 516
 |||||
 DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 |||||
 QY 517 TCCTCTCTCTCTGATCCCATCAGAGGTCATCAACCTCATTCGAACCAAGAAACTGTG 576
 |||||
 DB 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
 |||||
 QY 577 TCCTCAAAAGTGAGACATCCGCTGATCCCGTGAAGCTCTCTGATGAGCCCTC 636
 |||||
 DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProGluGluSerLeu 180
 |||||
 QY 637 ACTTGGCAGTATGGATCAGTTTGTGCGAATCTGGGGCTGCGAGGACGCTGGGC 696
 |||||
 DB 181 LysTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlyGlyLeuGly 200
 |||||
 QY 697 TCCCTTGGAAATCGGGAACAAAGGAGAAGAGTCTTCATCAGCTGTGAGGCTCCCGA 756
 |||||
 DB 201 SerProGlyAsnArgThrThrLysGluLysLysValPheIleSerLeuValGlySerArg 220
 |||||
 QY 757 GGCTTGGCTGAGCATTTCCAGCGGCCCATCCAGAGCCTGGCATCTTTATCAGCCAT 816
 |||||
 DB 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheValSerHis 240
 |||||
 QY 817 GTCAACCTGGCTCTCTGCTGAGTGGGATTCGATAGGACAGCATTCGCGAA 876
 |||||
 DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluThrGlyAspGlnIleValGlu 260
 |||||
 QY 877 GTCAATGGCTCGCATCTCTAACCTGGATCAAGAGGCTGTAAATGCTGTGAAAT 936
 |||||
 DB 261 ValAsnGlyIleAspPheThrAsnLeuAspHisLysGluAlaValAsnValLeuLysSer 280
 |||||
 QY 937 AGCCGAGCGCTGACCATCTCCATTTGATGCTGCGCGGAGCTGTTCATGACAGAC 996
 |||||

DB 281 SerArgSerLeuThrIleSerIleValAlaGlyAlaGlyArgGluLeuPheMetThrAsp 300
 |||||
 QY 997 CGGAGCGGCTGGCAGAGCGCGCGAGCTGAGCTGCAGCGGAGGAGTCTCTCATCGAC 1056
 |||||
 DB 301 ArgGluArgLeuGluGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuLeuMetGln 320
 |||||
 QY 1057 AAGCGGCTGGCGATGGAGTCCAAACAAGATCCTCCAGAGCAGCAGAGATGGAGCGCAA 1116
 |||||
 DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnGlnGlnGlnGln 340
 |||||
 QY 1117 AGGAGAAAGAAATTCGCCAGAGGAGCAGAGAGAAATGAGAGATACCCGAAAGGAGATG 1176
 |||||
 DB 341 ArgArgLysGluIleAlaGlnLysAlaIleGluAsnGluArgTyrArgLysGluMet 360
 |||||
 QY 1177 GAACAGATTCTAGAGAGGAGAGAAAGTTTAAAGAACAATGGGAAGAACTGGGGCTCA 1236
 |||||
 DB 361 GluGlnIleSerGluGluGluGluLysPheLysLysGlnTrpGluGluAspTyrGlySer 380
 |||||
 QY 1237 AAGGAACAGCTACTCTTGGCTTAAACCATCCTCCTGCTGAGGTACACCCAGTACCCCTTCGC 1296
 |||||
 DB 381 LysGluGlnLeuIleLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 |||||
 QY 1297 AAGCCA----- 1302
 |||||
 DB 401 LysProLysTyrAspGlnGlyValGluProAlaAspHisLeuAspGlySerThrGluGlu 420
 |||||
 QY 1303 -----AAGTGATTCGCGAAATATAGGAAGGCTTTGACCCCTACTCTATGTTTAC 1353
 |||||
 DB 421 GlnArgGlnGln-AspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheSe 440
 |||||
 QY 1354 CCCAGACAGATCATGGGAGGATGTCCGCTCCTACGATCAAGAAGGAGGATCCTT 1413
 |||||
 DB 440 rProGluGlnIleAlaGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLe 460
 |||||
 QY 1414 AGACTGGCTGCGAAGCGGTGGACTCCCGCATTTGGGAAGTGGTCTTCTGCTGT 1473
 |||||
 DB 460 uAspLeuAlaLeuGluGlyValAspSerProValGlyLysValValValSerAlaVal 480
 |||||
 QY 1474 GTATGAGCGGGAGCTGCTGAGCGGCATGTGTGGCATTTGTGAAGGGACGAGATCATGCG 1533
 |||||
 DB 480 lTyrGluGlyGlyAlaAlaGluArgHisGlyGlyValValLysGlyAspGluIleMetal 500
 |||||
 QY 1534 AATCAAGCGCAAGATTGTGACAGACTACCCCTGGCTGAGGCTGAGCTGCCCTGCAGAA 1593
 |||||
 DB 500 aileAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaGluAlaLeuGlnLys 520
 |||||
 QY 1594 GGCCTGGAATCAGCGCGGAGTGCATCGACTTGTGTTGCTGCTGCCCTGCCCCCAAGGA 1653
 |||||
 DB 520 salatrPasnGlnGlyGlyAspTrpIleAspLeuValValAlaValCysProLysGln 540
 |||||
 QY 1654 GTATGACCATGAGCTGACCTTC 1675
 |||||
 DB 540 uTyrAspAspGluLeuThrPhe 547
 |||||

RESULT 2

Q91XD1 PRELIMINARY; PRT; 548 AA.
 AC Q91XD1;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE RIKEN CDNA 2010016F01 gene.
 GN 2010016F01RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COLON;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BC010819; AAI10819.1; -.

DR MGD; MG1:1919338; 2010016F01R1k.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR PROSITE; PS50106; PDZ; 3.
SQ SEQUENCE 548 AA; 61711 MW; B00C4F00920ED76B CRC64;

Alignment Scores:
Pred. No.: 6.64e-177 Length: 548
Score: 2554.50 Matches: 502
Percent Similarity: 93.98% Conservative: 13
Best Local Similarity: 91.61% Mismatches: 11
Query Match: 64.04% Indels: 22
DB: 11 Gaps: 1

US-09-502-945-5 (1-2162) x Q91XD1 (1-548)

QY 97 ATGACCGAAGAGTGGCCGAGAAATTCGGCATATAGGTGGATTTCGATTGAAATGAT 156
|||||
Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuLeuGluAsnAsp 20
|||||

QY 157 GCAGAGAAGGACTATCTCTATGATGTGCTGGGAATGTACCACAGACCATGGAGCTGGCC 216
|||||
Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
|||||

QY 217 GTGCTCTGGGAGACCTGAAGCTGGTCAATCAATGAACCCAGCCCTCTGCTTGTGAT 276
|||||
Db 41 ValLeuValGlyAspLeuLysLeuValLeuAsnGluProAsnArgLeuProLeuPheAsp 60
|||||

QY 277 GCCATTGGCGCGTATCCCACTGAAGCACCAGGTGAATATGATGACCTGACCCCGGG 336
|||||
Db 61 AlaLeuArgProLeuLeuProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
|||||

QY 337 CGTCTCCAGGAAGTGAAGAGGTCGCTGTGACCGCTGTGACCCCGAGGCCCTCGGCGTG 396
|||||
Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
|||||

QY 397 AGTGTGGTGGTGGCCGTGGAGTTFGGTGTGGCTCTTCATCTCCCACTCATCAAGGC 456
|||||
Db 101 SerValArgGlyLeuGluPheGlyCysGlyLeuPheLeuSerHisLeuLeuLysGly 120
|||||

QY 457 GGTGAGCAGACAGCGTGGGCTCCAGGTAGGGAGCAGATCTGCTCCGGATCAATGGATAT 516
|||||
Db 121 GlyClnAlaAspSerValGlyLeuGlnValGlyAspGluLeuValArgIleAsnGlyTyr 140
|||||

QY 517 TCCATCTCCCTCTACCCATGAGGAGTGCATCAACCTCATTCGAACCAAGAAACGTGTG 576
|||||
Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuLeuArgThrLysLysThrVal 160
|||||

QY 577 TCCATCAAGTGAGACATCGGCGCTGATCCCGCTGAAAGCTCTCTGATGAGCCCTC 636
|||||
Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProGluGluSerLeu 180
|||||

QY 637 ACTTGGCAGTATGTGGATCAGTTGTCTCGAATCTGGGCGGTGGCAGCAGCCTGGGC 696
|||||
Db 181 LysTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlyLeuGly 200
|||||

QY 697 TCCCTCGAATCGGAAAAACAAGGAGAAGAGTCTTCATCAGCCCTGGTAGGCTCCCGA 756
|||||
Db 201 SerProGlyAsnArgThrThrLysGluLysValPheIleSerLeuValGlySerArg 220
|||||

QY 757 GGCCCTGGCTGCAGCATTTCCAGGGGCCCATCCAGAAAGCCTGCATCTTTATCAGCCAT 816
|||||
Db 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheValSerHis 240
|||||

QY 817 GTGAACCTGGCTCCCTGTCTGCTGAGTGGGATTTGAGATAGGGACAGATTTGCGAA 876
|||||
Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluThrGlyAspGlnIleValGlu 260
|||||

QY 877 GTCAATGGCTCGACTTCTCTAACCCTGGATCAACAAGGAGGCTGTAAATGTGCTGAAAT 936
|||||
Db 261 ValAsnGlyIleAspPheThrAsnLeuAspHisLysGluAlaValAsnValLeuLysSer 280
|||||

QY 937 AGCCGAGCCTGACCATCTCCATTGTAGCTGCAGCTGGCGGGAGCTGTTTCATGACAGAC 996
|||||

Db 281 SerArgSerLeuThrIleSerIleValAlaGlyAlaGlyArgGluLeuPheMetThrAsp 300
|||||

QY 997 CGGAGCGGCTGGCAGAGCGCGCAGCTGAGCTGCAGCGGAGGAGCTTCTCATGCAG 1056
|||||
Db 301 ArgGluArgLeuGluGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
|||||

QY 1057 AAGCGGCTGGCGATGGAGTCCACACAGATCTCCAGAGCAGCAGGAGATGGAGCGGCAA 1116
|||||
Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnGlnGlnGln 340
|||||

QY 1117 AGGAGAAAACAATTCGCCAGAGGCGCAGAGGAGAAATCAGAGATACCCGAGGAGATG 1176
|||||
Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
|||||

QY 1177 GAACAGATTCTAGAGGAGGAGAGAGTAAAGCAATATGGAAGCAACTGGGACTGGGCTCA 1236
|||||
Db 361 GluGlnIleSerGluGluGluGluLysPheLysLysGlnTrpLysGluAspTrpGlySer 380
|||||

QY 1237 AAGAACAGCTACTCTTCCTTAAACCATCTGCTGAGGTACACCCAGTACCCCTTCGC 1296
|||||
Db 381 LysGluGlnLeuIleLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
|||||

QY 1297 AAGCCA----- 1302
|||||

Db 401 LysProLysTyrAspGlnGlyValGluProAlaAspHisLeuAspGlySerThrGluGlu 420
|||||

QY 1303 -----AAGTGATTTCGGAAATATAGGAAGGCTTTGACCCCTACTCTATGTTAC 1353
::: |||||

Db 421 GlnArgGlnGln-AspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPhe 440
|||||

QY 1354 CCAGAGCAGATCATGGGAGGAGTCCGGCTCTACGCATCAAGAGGAGGATCCTT 1413
:|||||

Db 440 rProGluGlnIleAlaGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLe 460
|||||

QY 1414 AGACCTGGCCCTCGAAGCGGCTGTGGACTCCCGCATTTGGGAAGGTGCTCTTTCTGCTGT 1473
|||||

Db 460 uAspLeuAlaLeuGluGlyValAspSerProValGlyLysValValSerAlaVal 480
|||||

QY 1474 GTATGAGCGGGAGCTGTGAGCGGATGGTGCATTTGTAAGGGGAGCAGATCATGCC 1533
|||||

Db 480 lTyrGluGlyGlyAlaAlaGluArgHisGlyGlyValValLysGlyAspGluIleMetAl 500
|||||

QY 1534 AATCAAGCGCAACATGTGACAGACTACACCTGGCTGAGGCTGAGCTGCCCTGCAGAA 1593
|||||

Db 500 aIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaGluAlaLeuGlnLys 520
|||||

QY 1594 GGCTGGAATCAGGGCGGGGACTGGATCGACCTTGTGGTTGGCTGTGCCCTGCCCAAGGA 1653
|||||

Db 520 sAlaTrpAsnGlnGlyAspTrpIleAspLeuValValAlaValCysProProLysGlu 540
|||||

QY 1654 GTATGAGCATGACCTGACCTTC 1675
|||||

Db 540 uTyrAspAspGluLeuThrPhe 547
|||||

RESULT 3
Q96B29 PRELIMINARY; PRT; 533 AA.
AC Q96B29;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to PDZ-73 protein.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016057; AAL16057.1; -.

R^T the inner ear sensory hair cells, underlies Usher syndrome type 1C.;
 RL Nat. Genet. 26:51-55(2000).
 DR ENBL; AF228925; AAG12458.1; -
 DR HSP; P29476; IQAV.
 DR MGD; MGI:1919338; 2010016F0IRik.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR002965; P-rich_extensn.
 DR Pfam; PF00595; PDZ; 3.
 DR PRINTS; PRO1217; PRICHEXTENS.
 DR SMART; SM00228; PDZ; 3.
 DR PROSITE; PS0106; PDZ; 3.
 SQ SEQUENCE 910 AA; 102284 MW; 427B97953BA5D941 CRC64;

Alignment Scores:
 Pred. No.: 5,37e-166 Length: 910
 Score: 2405.50 Matches: 524
 Percent Similarity: 58.06% Conservative: 20
 Best Local Similarity: 55.92% Mismatches: 39
 Query Match: 354
 Indels: 7
 DB: 11
 Gaps: 7

US-09-502-945-5 (1-2162) x Q9ES64 (1-910)

QY 97 ATGCACGAAAGTGGCCGAGATTCGCCGATTAAGTGGATTCTCTGATTGAAATGAT 156
 Db 1 MetaspargLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
 QY 157 GCAGAGAAGGACTATCTCTATGATGTGCTCGGAATGTACCACAGACCCATGGACGTGGCC 216
 Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetaspValAla 40
 QY 217 GTCTCGTGGAGACCTGAAGCTGGTCAATGAACCCAGCGCTGCGCTCTGTTTGTAT 276
 Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProAsnArgLeuProLeuPheAsp 60
 QY 277 GCATTGCGCGCTGATCCCACTGAAGCACCAGCTGGAATATGATCAGCTGACCCCGCG 336
 Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 QY 337 CGCTCCAGGAAGCTGAAGAGGTGCTGTCGACCGCTGCACCCGAGCGCTCGGCCCTG 396
 Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
 QY 397 AGTGTGCTGGTGGCTCGGATTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGGC 456
 Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
 QY 457 GTCAGGCAGACAGCTCGGCTCCAGGTAGGGACGACATGCTCCGGATCAATGATAT 516
 Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 QY 517 TCCATCTCTCTGTACCCATCAGGAGGTCAATCAACCTCATTCGACCAAGAAACTGTG 576
 Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
 QY 577 TCCATCAAGTGAACATCGGCTGTATCCCGTGAAGAGCTCTCTCATGAGCCCTC 636
 Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProGluGluSerLeu 180
 QY 637 ACTTGCCAGTATGTGATCAGTTTGTGTCGGAATCTGGGGCGTGCAGGACGCTGGGC 696
 Db 181 LysTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlyLeuGly 200
 QY 697 TCCCTTGGAAATCGGAAACAGAGAGAGAGTCTTCATCAGCTGGTAGGCTCCCGA 756
 Db 201 SerProGlyAsnArgThrThrLysGluLysLysValPheIleSerLeuValGlySerArg 220
 QY 757 GGCCTTGGCTGCGCATTTCCAGCGCCCATCCAGAACCTGGCATCTTTATCAGCCAT 816
 Db 221 GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheValSerHis 240
 QY 817 GTGAACCTGGCTCCCTGCTGCTGAGGTGGATTTGGAGATAGGGGACCATGCTCGAA 876
 Db 1 MetaspargLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20

Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluThrGlyAspGlnIleValGlu 260
 QY 877 GTCATGGCGTGCATCTCTTAACCTGGATCAACAAGAGGCTGTAATGTGCTGAAAAAT 936
 Db 261 ValAsnGlyIleAspPheThrAsnLeuAspHisLysGluAlaValAsnValLeuLysSer 280
 QY 937 AGCCGACGCTGACCATCTCCATCTAGCTGACGTGGCCGGAGCTGTTTCATGACAGAC 996
 Db 281 SerArgSerLeuThrIleSerIleValAlaGlyAlaGlyArgGluLeuPheMetThrAsp 300
 QY 997 CGGAGCGCTGCGACAGCGCGGCGAGCTGAGCTGCAGCGGAGGAGTCTTCATGTCAG 1056
 Db 301 ArgGluArgLeuGluGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
 QY 1057 AACCGCTGCGATGAGTCCCAACAGATCTCCAGAGACAGCAGAGATGGAGCGCAA 1116
 Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGluGlnMetGluArgGln 340
 QY 1117 AGGAGAAAATAATGCCAGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1176
 Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
 QY 1177 GAACAGATTGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1236
 Db 361 GluGlnIleSerGluGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380
 QY 1237 AGGAGACACTACTCTTGGCTAAAACCATCTACTGCTGAGGTACACCCAGTACCCCTTGC 1296
 Db 381 LysGluGlnLeuIleLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 QY 1297 AAGCCA----- 1302
 Db 401 LysProLysSerPheGlyTrpPheTyrArgTyrAspGlyLysPheProThrIleArgLys 420
 QY 1302 ----- 1302
 Db 421 LysAlaLysGluLysLysLysAlaLysTyrAspSerLeuGlnAspLeuArgLysAsnLys 440
 QY 1302 ----- 1302
 Db 441 LysGluLeuGluPheGluGlnLysLeuTyrLysGluLysGluGluMetLeuLysGlu 460
 QY 1302 ----- 1302
 Db 461 LysGlnLeuLysIleAsnArgLeuAlaGlnGluValSerGluThrGluArgGluAspLeu 480
 QY 1302 ----- 1302
 Db 481 GluGluSerGluLysThrGlnTyrTrpValGluArgLeuCysGlnThrArgLeuGluGln 500
 QY 1302 ----- 1302
 Db 501 IleSerSerAlaGluAsnGluIleProGluMetThrThrGlyProProProProProPro 520
 QY 1302 ----- 1302
 Db 521 SerValSerProLeuAlaProProLeuArgArgPheAlaGlyGlyIleHisLeuHisThr 540
 QY 1302 ----- 1302
 Db 541 ThrAspLeuAspAspIleProLeuAspMetPheTyrTyrProProLysThrProSerAla 560
 QY 1302 ----- 1302
 Db 561 LeuProValMetProHisProProSerValAsnSerProSerLysValProAlaProPro 580
 QY 1302 ----- 1302
 Db 581 ValLeuProSerSerGlyHisValSerSerSerSerProTrpValGlnArgThrPro 600
 QY 1302 ----- 1302
 Db 601 ProProIleProIleProProProProSerIleProThrGlnAspLeuThrProThrArg 620

```
QY 1302 ----- 1302
Db 621 ProLeuProSerAlaLeuGluAlaLeuGlyAsnHisProPheArgThrGlyAspPro 640
QY 1302 ----- 1302
Db 641 GlyHisProAlaAspAspTrpGluAlaAsnThrHisSerGlyLysProSerSerPro 660
QY 1302 ----- 1302
Db 661 ThrThrGluArgSerPheProProAlaProLysThrPheCysProSerProGlnProPro 680
QY 1302 ----- 1302
Db 681 ArgGlyProGlyValSerThrIleSerLysProValMetValHisGlnGluHisAsnPhe 700
QY 1302 ----- 1302
Db 701 ValTyrArgProAlaValLysSerGluValLeuProGlnGluMetLeuLysArgMetVal 720
QY 1303 -----AAGTGATTCCGGAAATATCATCAGGAAGCTTTGACCCCTA 1341
Db 721 ValTyrGlnThrAlaPheArgGln-AspPheArgLysTyrGluGlyPheAspProTy 740
QY 1342 CTCATGTTACCCAGAGCATCATCGGGAAGGATGTCCTCGGCTCTACCATCAAGAA 1401
Db 740 rSerMetPheSerProGluGlnIleAlaGlyLysAspValArgLeuLeuArgIleLysLy 760
QY 1402 GGAGGATCTTAGACCTGGCCCTGGGAAGCGGTGTGGACTCCCTCCATTTGGGAAGTGGT 1461
Db 760 sGluGlySerLeuAspLeuAlaLeuGluGlyGlyValAspSerProValGlyLysValVa 780
QY 1462 CGTTCTCTGTTGATGAGCGGGAGCTGCTGAGCGGATGTCGTCATGTGGAAGGGGA 1521
Db 780 lValSerAlaValTyrGluGlyGlyAlaAlaGluArgHisGlyGlyValValLysGlyAs 800
QY 1522 CGAGATCATGCCAATCAACGCGCAAGATGTGTACAGACTACACCTCGGCTGAGGCTGACG 1581
Db 800 pGluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaGluAl 820
QY 1582 TGCCCTGAGAGGCTGGAATCAGGCGGGGAGTGGATGCACTTGTGTTGGCTGCTG 1641
Db 820 aAlaLeuGlnLysAlaTrpAsnGlnGlyGlyAspTrpIleAspLeuValAlaValAlcY 840
QY 1642 CCCCCCAAGGAGTAGTACGATGAGTACCTCTGTGAAGTCCAAAGGGGAACCA 1701
Db 840 sProProLysGluTyrAspAspGluLeuSerSerLeuProSerAlaAlaGluSer-- 859
QY 1702 AATTCACGCGTTAGGAAACAGTGAAGTCCGCGCCACCTCGTGAACACAAAGCCTCGGAC 1761
Db 860 -----ProGlnLeuAla-----ArgLy 865
QY 1762 CAGCCTTGAGAGGCGCCACATCACACACAGATGGATCTTTGGGACCTCAATCTATC 1821
Db 865 sGlnLeuGlu---AlaTyrGluProValCysArg---HisGlyPhePheLeuGlnLeuG 883
QY 1822 ACCCAGGAATCTCAAACTCCCTTTGGCCCTGAACACAGGCGCCAGATAAGAACAGCTCGG 1881
Db 883 uProThrAsnLeuLeuLeu-----LysSerArgGluArgAsnGlnThr-- 897
QY 1882 CCACCTTTTGAAGCCAAATGTGGAGAAAGGAGAGCAGCCGCG 1926
Db 898 -----Asp--ProSerTrpArgProAlaSerSerAlaPro 908
RESULT 5
Q9CVG7 PRELIMINARY; PRT; 112 AA.
AC Q9CVG7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 2010016F01RIK protein (Fragment).
```

```
GN 2010016F01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata Y., Konno H., Adachi J., Fukuda S.,
RA Arakawa T., Hara A., Fukunishi Y., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Hoffell D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK008274; BAB25568.1; -
DR HSSP: Q12923; 3PDZ; 2010016F01RIK.
DR MGD; MGI:1919338; 2010016F01RIK.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER
SQ SEQUENCE 112 AA; 11965 MW; BE42F80C7B3AB196 CRC64;

Alignment Scores:
Pred. No.: 1.6e-31 Length: 112
Score: 548.00 Matches: 105
Percent Similarity: 98.20% Conservative: 4
Best Local Similarity: 94.59% Mismatches: 2
Query Match: 13.74% Indels: 0
DB: 11 Gaps: 0

US-09-502-945-5 (1-2162) x Q9CVG7 (1-112)
QY 1343 TCATGTTCCACCCAGCAGATCATGGGAAGGATGTCGGCTCTTACCATCAAGAAG 1402
Db 1 SerMetPheSerProGluGlnIleAlaGlyLysAspValArgLeuLeuArgIleLysLys 20
QY 1403 GAGGATCTCTTAGACCTGGCCCTGGGAAGCGGTGTGGACTCCCTCCATTTGGGAAGTGGTC 1462
Db 21 GluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProValGlyLysValVal 40
QY 1463 GTTCTCTGCTGTATGAGCGGGAGCTGCTGAGCGCATGTCGTCATTTGTAAGGCGAC 1522
Db 41 ValSerAlaValTyrGluGlyGlyAlaAlaGluArgHisGlyGlyValValLysGlyAsp 60
QY 1523 GAGATCATGCAATCAACGCGGATTTGTACAGACTACACCTCGGCTGAGGCTGACGCT 1582
Db 61 GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaGluAla 80
QY 1583 GCCCTGCAGAGGCTCGAATCAGGCGGGGAGCTGGGATCGACCTTGTGGTGGCTGCTGC 1642
Db 81 AlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValAlaValLys 100
QY 1643 CCCCCAAGGAGTAGTACGATGAGCTGACCTTCACCTTC 1675
Db 101 ProProLysGluTyrAspAspGluLeuThrPhe 111
```

```

RESULT 6
Q9W443 PRELIMINARY; PRT; 493 AA.
AC Q9W443;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG5921 protein.
GN CG5921.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Artil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler J., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K., Nixon K., Nusskern D.R., Pacleib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003436; AAF46116.1; .
DR HSSP; Q12959; 1pDR.
DR FlyBase; FBgn0029835; CG5921.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
SQ SEQUENCE 493 AA; 50574 MW; C46711C9F19797A4 CRC64;

Alignment Scores:
Pred. No.: 8,06e-22
Score: 418.00
Percent Similarity: 45.89%
Best Local Similarity: 30.92%
Query Match: 10.48%
DB: 5
Length: 493
Matches: 128
Conservative: 62
Mismatches: 157
Indels: 67
Gaps: 12

```

```

US-09-502-945-5 (1-2162) x Q9W443 (1-493)
QY 340 TCCAGGAAGCTGAAGAGGCTGCTGACCGCTGTGCACCCGAGGCCCTC----- 390
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 58 SerSerArgLeuArgValLeuArgLeuValArgProHisGlnArgArgLeuLeuValGly 77
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 391 -----GGCCTGAGTGTGCTGGTGGCTGGAGTTGGCTGT 426
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 78 GlyProGluArgGlySerThrThrGlyPheThrValArgGlyGlyArgGluHisGlyThr 97
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 427 GGGCTCTTCATCTCCACCTCATCAAGAGCGGTGACGACAGCGTGGCTCCAGGTA 486
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 98 GlyPhePheValSerHisValGluHisGlyGlyGluAlaHisLeuLeuGlyLeuArgIle 117
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 487 GGGACGAGATCGTCCGATCAATGGATATTCATCTCTCTGTTACCCATGAGGAGTC 546
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 118 GlyAspGlnIleLeuArgIleAsnGlyPheArgLeuAspAlaValHisGlyGluPhe 137
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 547 ATCAACCTCATTCGAACCAAGAAACTGTCTCCATCAAGTGAACATCGCCCTCATC 606
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 138 IleGlnLeuValAlaGlyGlnAspArgValThrLeuLysValArgGlyValGlyMetLeu 157
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 607 CCCGTGAAAGCTCTCTCTGATGAGCCCTCCTTGGCAGTATGTGGATCATAGTTGTGTCG 666
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 158 ProValArgAspLeuProGluGluArgLeuSerTrpSerValValLys-----LeuPro 175
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 667 GAATCTGGGGGTGGCGAGCGCTGGGCTCCCTGGGAATCGGAAAAACAAGGAGAAG 726
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 176 SerValSerGlyThrProSerGluSerSerPheLysGlyGluArgArgGlyAlaSerArg 195
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 727 AAGCTTCTTCATCAGCTG---GTAGGCTCCGA---GGCTTGGCTGCACCATTTCCAGC 780
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 196 AspIleSerValValLeuHisValAlaProArgThrLysLeuGlyLeuGlyIleCysLys 215
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 781 GGCCCCATCCAGAGCTGGCATCTTTATCAGCATGTGAAACCTGGCTCCCTGTCTGCT 840
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 216 GlyProGluTrpLysProGlyIlePheValGlnPheThrLysAspArgSerValAlaArg 235
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 841 GAGTGGGATTGAGATAGGGGACAGATGTGCAAGTCAATGGCTGCGACTTCTTAAC 900
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 236 GluAlaGlyLeuArgProGlyAspGlnIleLeuSerValAsnSerIleAspPheSerAsp 255
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 901 CTGGATCACAAAGAGCTGTAAATGTGCTGAAATAAGCGCGCTGACCATCTCCAT 960
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 256 ValLeuPheSerGluAlaValAlaValMetLysSerSerLysLeuAspMetValVal 275
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 961 GTAGCTGACGTGGCGGAGCTGTTTCATGACAGACCGGCGCTGGCAGAGCGCGG 1020
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 276 ArgThrAlaAlaGlyCysAspLeuPheProGlyGlu-SerSerGlyTyrAsnSerSerAl 295
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 1021 CAGCGTGAGCTGCGCGGAGGAGCTTCTCATGCAAGACGCGCTGGCGATGAGTCCAAC 1080
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 295 aSer---SerValThrGlyAspGlnSerProCys-----TIPAlaAspAl 309
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 1081 AAGATCTCCAGGACGAGGAGGATGGAGCGGCAAGAGAAAAGAAATTCGCCAGAAG 1140
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 309 alySerLysArg-LeuThr-----AlaValArgGluGluSerGlyAlaGlyGlyG 326
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 1141 GCAGCAGAGGAAAATGAGAGATACCGGAAGAGAGATGGAACAGATTGTAGAGGAGAGAG 1200
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 326 lyGlyCysGlyLeuSerSerAlaProGlyAlaGlySerProAsnTrpSerGlnGlyValG 346
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 1201 AGTTTAAAGCAATGGAAGAAAGACTG-----GGCTCAAGAGAA 1242
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 346 luValHisLysGlnMetAsnLysThrIleIleLysLeuThrGluAsnGlyThrSerIleA 366
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 1230 -----GGCTCAAGAGAA 1242
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 366 snAsnThrTyrIleAlaSerThrGlyGlySerValSerGlySerGlyThrGlyS 386
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
QY 1243 CAGCTACTCTTGGCTAAAACCATCACTGTGAGGTACACCCCTCCGCAAGCCA 1302
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 386 erGlyThrSerGlyArgSerGlnGlnSerGlnSerAsnProSerArgAsn- 405
   ||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|

```

```
QY 1303 AAGTATTTCGGAAATATGAGGAGCT---TTGACCCCTACT------ 1343
D 406 -----SerThrThrMetLysArgSerHisLeuArgProValAsnSerAlaGlySerG 423
QY 1344 -----CTATGTTTACCCAGAGACAGATCATGGGGAAGG-----ATGTCCGGCTCT 1389
D 423 lYlleglyLeuSerSerGlySerAlaGlySerAlaGlySerAlaGlySerSerGlySerG 443
QY 1390 ACGCATCAAGAGGAGGATCTTAGACTGGCCCT 1425
D 443 lYSer-ArgSerGlyGlyValileAlaProAlaPro 454
RESULT 7
Q8WVS2
ID Q8WVS2 PRELIMINARY; PRT; 906 AA.
AC Q8WVS2;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 96.4 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARY;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014524; AAH14524.1;
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS50106; PDZ; 3.
KW Hypothetical protein.
SQ SEQUENCE 906 AA; 96428 MW; D8D75D87A41A5E05 CRC64;

Alignment Scores:
Pred. No.: 9, 11e-21 Length: 906
Score: 405.00 Matches: 196
Percent Similarity: 36.79% Conservative: 102
Best Local Similarity: 24.20% Mismatches: 285
Query Match: 10.15% Indels: 227
DB: 4 Gaps: 30

US-09-502-945-5 (1-2162) x Q8WVS2 (1-906)
QY 7 CCGGTCCGGTCCGGCTCTTTCCAGCTCTCGCAGCCGGGACCCGGAAGGACGGTCTG 66
D 4 ProLeuAspGlyLeuSerValSerSer-----SerThrGlySerLeuGlySer 20
QY 67 TGCACGACGACGCTGGACCTGGCCCGCCATGGACCGAAAGTGGCCCGGAGAAATTCGG 126
D 21 AlaAlaGlyAlaGlyGlyGlyAlaGlyGlyAlaGlyLeuArgLeuLeuSerAlaAsnValArg 40
QY 127 -----CATAAAGTGGATTTCTGATGAAATGATGACAGAAAGGACTATCTCTATGAT 180
D 41 GlnLeuHisGlnAlaLeuThrAlaLeuLeuSerGluAlaGluArgGluGlnPheThrHis 60
QY 181 GTCTCGGAATGTACACACGACCATGGAGCTGGCGCTGCTCTGGGAGACCTGGAAGCTG 240
D 61 CysLeuAsnAlaTyrHisAlaArgArgAsnValPheAspLeuValArgThrLeuArgVal 80
QY 241 GTCATCAATGAACCCAGCCGCTCGCTCTGTTGATGCCATTCGGCCGCTGATCCCACTG 300
D 81 LeuLeuAspSerProValLysArgArgLeuLeuProMetLeuArgLeuValIleProArg 100
QY 301 AACCAACGCTGGAATATGATCAG-----CTGACC 330
D 101 SerAspGlnLeuLeuPheAspGlnTyrThrAlaGluGlyLeuTyrLeuProAlaThrThr 120
QY 331 CCGCGGCGC-----TCCAGGAAGCTGAAGGAGGTGCGTCTG 366
```

```
Db 121 ProTyrArgGlnProAlaTrpGlyGlyProAspSerAlaGlyProGlyGluValArgLeu 140
QY 367 GACCGTCTGCAC-----CCGGAAGGCTCGGCTGAGTGTGCTGTGGCTG 414
D 141 ValSerLeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySer 160
QY 415 GAGTTTGGCTGGGCTCTTCATCTCCACCTCATCAAGGCGGTTCAGGACAGACGGTC 474
D 161 GluHisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGlu 180
QY 475 GGGTCCAGGTAGGGACAGATCGCTCCGATCAATGATATCCATCTCTCTCTGACC 534
D 181 GlyLeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThr 200
QY 535 CATGAGGAGGTCAATCAACCTCATTCGAACCAAGAAACTGTGTCCATCAAGTGCAGAC 594
D 201 HisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSer 220
QY 595 ATCGGCTGATCCCGGTGAAAGCTCTCTCGATGAGCCCTCACTGGCAGTATGTGAT 654
D 221 AlaGlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAsp 238
QY 655 -----CAGTTGTGTCGGAATCTGGGGCTGGAGGCGAGCTGGG----- 696
D 239 ProGlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeu 258
QY 697 -----TCCCTGGAAATCGGGAAG-----AAGGAGAAG 726
D 259 ArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLys 278
QY 727 AAGTCTTTCATCAGCCTGGTAGGTCGCGAGGCTTGGCTGCAGCATTTCAGCGGCCCC 786
D 279 LysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAla 298
QY 787 ATCCAGAAGCTGGCATCTTTATCAGCCATCTGAACCTGCTCCTCTCTCTGCTGAGGTG 846
D 299 GluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySer 318
QY 847 GGATTGGAGATAGGGACGACGATTCGAAAGTCAATGCGCTCGACTTCTTAACCTGGAT 906
D 319 GlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeu 338
QY 907 CACAAGGAGGTGTAAATGTCTGAAATAGCCGACGCTGACCATCTCCTCTCTGCTACT 966
D 339 HisAspGluAlaValArgLeuLysSerSerArgHisLeuLeuLeuThr---ValLys 357
QY 967 GCAGCTGCCCG-----GAGCTGTCATGACAGACGCGGAGCGGTGCGACAG 1014
D 358 AspValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSer 377
QY 1015 GCGCGGACGCTGAGCTGCAGCGGAGGAGCTTCTCATGCAAGCGGCTGGCGATGAG 1074
D 378 SerArgIleArgGluThrMetAlaAsnSerAlaGlyPheLeuGlyAspLeuThrThrGlu 397
QY 1074 ----- 1074
D 398 GlyIleAsnLysProGlyPheTyrLysGlyProAlaGlySerGlnValThrLeuSerSer 417
QY 1075 -----TCCAAACAGATCCTCCAGGAGCAGCAGAGATGGAGCGCAAGAGGAAAA 1125
D 418 LeuGlyAsnGlnThrArgValLeuLeuGluGluGlnAlaArgHisLeuLeuAsnGluGln 437
QY 1126 GAAATTGCCAGAGGACGAGGAGAAATGAGAGATACCGG----- 1167
D 438 GluHisAlaThrMetAlaTyrTyrLeuAspGlu---TyrArgGlyGlySerValSerVal 456
QY 1168 -----AAGGAGATGGACACGATTTCTAGAGGAGGAGAGAGATTTAAGAGCAA 1215
D 457 GluAlaLeuValMetAlaLeuPheLysLeuLeuAsnThrHisAlaLysPheSerLeuLeu 476
QY 1216 TGGGAA-----GAAGACTGGGCTCAAGGACACGACTACTCTTG 1254
D 1216 TGGGAA-----GAAGACTGGGCTCAAGGACACGACTACTCTTG 1254
```

```
Db 477 SerGluValArgGlyThrIleSerProGlnAspLeuGluArgPheAspHisLeuValLeu 496
QY 1255 CCTAAACCATCAGTCGTGAGTACACCCAGTACCCCTTCGCAAGCCAAAGTATTCCG 1314
Db 497 ArgArg-:::||||:||||:||||:||||:||||:||||:||||:||||:||||: 508
QY 1315 GAAATATGAGGAAGCTTTGACCCCTACTCTATGTATCCACCCAGACAGATCATGGGAA 1374
Db 508 ----- 508
QY 1375 GGATGTCGGCTCTACGCATCAAGAAGGAGGATCCTTACGCTGGCCCTGGAGCGG 1434
Db 509 -----ProGlyProGly----- 512
QY 1435 TGTGACTCCCCATTGGGAAGGTGCTTCTCTGTCTGTATGAGCGGGAGCTGCTGA 1494
Db 513 -----AlaGlyAspThrTyr 517
QY 1495 GCGGCATGTTGTCATTTGTAAGAGGGACGAGATCATGGAATCAA----- 1539
Db 518 SerMetValSerTyrSerAspThrGlySerSerThrGlySerHisGlyThrSerThrThr 537
QY 1540 -----CGCAAGATTGTACAGACTACACCTGCTGAGCTGAGCGTGCCT 1587
Db 538 ValSerSerAlaArgAsnThrLeuAspLeuGluThrGly----- 551
QY 1588 GCAGAAAGCCCTGGATCAGGCGGGAGCTGATCGACCTTGTGTTGCTGCTGCCCTCC 1647
Db 552 ---GluAlaValGlnGlyAsnIleAsn-----AlaLeuProAsp 563
QY 1648 AAAGGATGATGACGATGACCTTCTTGTGAAGTCCAAAGGG----- 1695
Db 564 ValSerVal-----AspValArgSerThrSerGlnGlyLeuSerSerPhe 579
QY 1696 AAACCAATTCACGGTTAGGAACAGTACCTCGGCCACCTCGTGAACAAAGCC 1755
Db 580 LysPro-----LeuProArgProProProLeuAlaGlnGly 591
QY 1756 TCGACACAGCTTGAGAGAGCCACATGACACACACAGATGGCATCTTGGGACCTGAA 1815
Db 592 AsnAspLeuProLeuGlyGlnProArg-----LysLeuGlyArgGlu 605
QY 1816 TCTATACCCAGGAATCTCAACTCCC----- 1842
Db 606 AspLeuGlnProProSerSerThrProSerCysSerGlyThrValPheSerAlaProGln 625
QY 1843 -----TTGGGCTGACACAGGCGCCAGATAGGACAGCTCGG 1881
Db 626 AsnArgSerProProAlaGlyThrAlaProThrProGlyThrSerSerAlaGlnAspLeu 645
QY 1882 CCACCTTTTGAAGGCCAATGTGAGGAAGGAGC---AGCCAGCGTTTGGGAGAAGA 1938
Db 646 ProSer-----SerProIleThrAlaSerValSerProAlaAsnProSerSerLysArg 663
QY 1939 TCTCAAGATCCAGACTCTCATCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1998
Db 664 ProLeuAspAlaHisLeuAlaLeuValAsnGlnHisProIle----- 677
QY 1999 TTTGGGGACTCTCTCTGACCTTAATAGACCCAGCTGAGTCTCTCTCTCTCTCT 2058
Db 678 -----GlyProPheProArgValGlnSerProProHis-----LeuLys-SerProSe 693
QY 2059 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2086
Db 693 rAlaGluAlaThrValAlaGlyGlyCys 702
RESULT 8
Q9P202
ID Q9P202 PRELIMINARY; PRT; 963 AA.
AC Q9P202
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

```
DE KIAA1526 protein (Fragment).
GN KIAA1526.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human
RT genes XVII. The complete sequences of 100 new cDNA clones from brain
RT which code for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
DR EMBL; AB040959; BAA96050.1; -.
DR HSP; P29476; IQAV.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PSS0106; PDZ; 3.
FT NON_TER 1
SQ SEQUENCE 963 AA; 102383 MW; 4D76D3A1CB052A2E CRC64;
```

```
Alignment Scores:
Pred. No.: 1,2e-20 Length: 963
Score: 403.50 Matches: 200
Percent Similarity: 37.97% Conservative: 103
Best Local Similarity: 25.06% Mismatches: 294
Query Match: 10.12% Indels: 201
DB: 4 Gaps: 28
```

US-09-502-945-5 (1-2162) x Q9P202 (1-963)

```
QY 7 CCGTCTCGGTCGCGTCTTCTCAGCTCTGCGAGCGGCGCACCGAAGGACGGGTCG 66
Db 60 ProLeuAspGlyLeuSerValSerSer-----SerThrGlySerLeuGlySer 76
QY 67 TGCACGACCGCAGCTGGACCTGCGCCAGCCATGGACCGAAAGTGGCCCGAGATCCCG 126
Db 77 AlaAlaGlyAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 96
QY 127 -----CATAGGTGGATTTCTGATTGAAATGATCCAGAGAGGACTATCTATGAT 180
Db 97 GlnLeuHisGlnAlaLeuThrAlaLeuLeuSerGluAlaGluArgGluGlnPheThrHis 116
QY 181 GTCTGCGAATGTACCACACGACCATGGAGTGGCGGCTGCTGCTGGGAGACCTGGAAGCTG 240
Db 117 CysLeuAsnAlaTyrHisAlaArgArgAsnValPheAspLeuValArgThrLeuArgVal 136
QY 241 GTCATCAATGAACCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Db 137 LeuLeuAspSerProValLysArgArgLeuLeuProMetLeuArgLeuValIleProArg 156
QY 301 AAGCACGAGTGGAAATATGATCAG-----CTGACC 330
Db 157 SerAspGlnLeuLeuPheAspGlnTyrThrAlaGluGlyLeuTyrLeuProAlaThrThr 176
QY 331 CCCGCGCGC-----TCCAGGAAGCTGAAGGAGGTGCGTCTG 366
Db 177 ProTyrArgGlnProAlaTrpGlyGlyProAspSerAlaGlyProGlyGluValArgLeu 196
QY 367 GACCGTCTGCAC-----CCCGAAGGCTCGGCTGAGTGTGGTGGTGGCTG 414
Db 197 ValSerLeuArgArgAlaLysAlaHisGluGlyGlyPheSerIleArgGlyGlySer 216
QY 415 GAGTTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGCGGTGAGCAGACACCGTC 474
Db 217 GluHisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGlu 236
QY 475 GGGCTCAGGTAGGAGGACGAGATCGTCCGATCAATGGATATTCATCTCTCTCTCTCT 534
Db 237 GlyLeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThr 256
```


[illegible]


```
Db 763 GlyAspLysLeuLeuGluValAsnGlyValAlaLeuGlnAspAlaGluHisGluAla 782
QY 547 ATCAACCTCATTCGA---ACCAAGAAACTGTGTCCTCAAGTG----- 588
Db 783 ValGluAlaLeuArgGlyAlaGlyAlaValGlnMetArgValTrpArgGluArgMet 802
QY 589 -----AGACATATGGCTGATCCCGCTGAAAGCTCTCCTGAT---GAGCCC 633
Db 803 ValGluProGluAsnAlaValThrIleThrProLeuArgProGluAspAspTyrSerPro 822
QY 634 CTCACCTGGCAGATGTGGATCAGTTGTGCGGAATCTGGGGCGTGGCAGCGAGCTG 693
Db 823 ArgGluTrpArg-----GlyGlyGlyLeuArgLeuProLeu 834
QY 694 GGCTCCCTCGAAATCGGGAACAAGGAGAAGAGTCTTCATCAGCCCTGGTAGGCTCC 753
Db 835 LeuGlnProGluThrProValSerLeuArgGlnArgHisAlaLaCysLeuValArgSer 854
QY 754 ---CGAGGCTTGGCTGCAGCATTTCCAGCGGCCCTCCAGAGCCT----- 798
Db 855 GluLysGlyLeuGlyPheSerIleAlaGlyLysGlySerThrProTyrArgAlaGly 874
QY 799 -----GGCATCTTATCAGCCATGTCGAAGTCAATGGCTGCTGCTGCTGAGTGGGA--- 849
Db 875 AspGlyIlePheIleSerArgIleAlaGluGlyAlaAlaHisArgAlaGlyThr 894
QY 850 TTGGAGATAGGACGACAGATGTGCAAGTCAATGGCTGCTGCTGCTGCTGCTGCTGCTG 909
Db 895 LeuGlnValGlyAspArgValLeuSerIleAsnGlyValAspMetThrGluAlaArgHis 914
QY 910 AAGAGGCTTAATGTGCTGGAATAATAGCCGAGCCTGACCATCTCCATCTGAGTGTGCA 969
Db 915 AspHisAlaValSerLeuLeu---ThrAlaAlaSerProThrIleSerLeuLeuLeuGI 933
QY 970 GCTGGCGGGAGCTGTCATGACAGACCGGCGGCTGCGAGAGGCGGCGAGCGTGAG 1029
Db 933 uArg-----GluThrGly----- 941
QY 1030 CTCAGCGCGCAGGAGCTTCTTCATGCGAGAGCGGCTGCGAGTGGAGTCCACAGATCCCTC 1089
Db 941 oProSerProProHisSer-----SerProThrProAlaAl 954
QY 1090 CAGGACGACGAGAGATGGAGCGGC-----AAAGGAGAAAGAAATGCCAGAGGCA 1143
Db 954 aThrValAlaAlaThrValSerThrAlaValProGlyGluProLeuLeuProArg----- 972
QY 1144 GCAGAGAAATAGAGATACCGGAGGAGATGGAACAGATGTAGAGGAGGAGAGAAG 1203
Db 972 ----- 972
QY 1204 TTTAAGAGCAATGGGAAGAAGACTGGGCTCAAGGAACAGCTACTCTTGCTTAAACCC 1263
Db 973 -----LeuSerPr 975
QY 1264 ATCAGTGTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATTTCGGGAATATGA 1323
Db 975 oSerLeuLeuAlaThrAlaLeu----- 982
QY 1324 GAGAGCTTTGACCCCTACTCTATGTTTACCCAGAGCAGATCATGCGGAAGGATGTCG 1383
Db 983 -GluGly-----ProTyrProVal-----GluGluIleCy 992
QY 1384 GCTCTTACCATCAAGAAGGAGGATCTTAGACCTGGCCCTGGAGGCGGTGTGAC--- 1441
Db 992 sLeuProArg-----AlaGlyGlyProLeuGlyLeuSerIleValGlyGlySerAspHI 1010
QY 1442 -----TCCCCCATTTGG-----AAGTGGTGGCTTTCTGCTGTGTATGAGCG 1482
Db 1010 sSerSerHisProPheGlyValGlnAspProGlyValPheIleSerLysValLeuProAr 1030
QY 1483 GGGAGCTGTGAGCGGATGTGGCATTTGTAAGGGGACGAGATCATGGCAATCAAGCG 1542
Db 1483 ----- 1542
```

```
Db 1030 gGlyLeuAlaAlaArg---CysGlyLeuArgValGlyAspArgIleLeuAlaValAsnGI 1049
QY 1543 CAGATTGTGACAGACTACACCTGCTGAGCTGAGCTGAGCTGAGCTGCTG 1588
Db 1049 yGlnAspValArgGluAlaThrHisGlnGluAlaValSerAlaLeu 1064
RESULT 12
075970 PRELIMINARY; PRT; 2042 AA.
ID 075970;
AC 075970;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Multi PDZ domain protein MUPP1.
GN MUPP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Eng L., Krapivinsky G., Clapham D.E.;
RT "Human homolog of MUPP1 protein."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF093419; RAC61870.1;
DR HSSP; Q12959; IPDR.
DR InterPro; IPR004172; L27.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF02828; L27; 1.
DR Pfam; PF00595; PDZ; 13.
DR SMART; SM00226; PDZ; 13.
DR PROSITE; PS0106; PDZ; 13.
SQ SEQUENCE 2042 AA; 218595 MW; F0E1DD029749A762 CRC64;
Alignment Scores:
Pred. No.: 4.88e-11 Length: 2042
Score: 273.00 Matches: 146
Percent Similarity: 36.78% Conservative: 71
Best Local Similarity: 24.75% Mismatches: 170
Query Match: 6.84% Indels: 203
DB: 4 Gaps: 25
US-09-502-945-5 (1-2162) x 075970 (1-2042)
QY 11 TCGCGTCCGGCTCTTTCCAGCTCTGCGACGCGGCGACCGAA-----GGAACG 61
Db 1585 SerGlnSerLeuMetValProGlnSerGlySerProGluProGluSerIleArgAsnThr 1604
QY 62 GGTCTGTGCAACGACGACGCTGGACCTGGCCCGACCATGGACCGAAAGTGGCCGAGAT 121
Db 1605 SerArgSerSerThrProAlaIlePheAla-Ser---AspProAlaThrCysProIleI 1623
QY 122 TCCGCAATAAGTGGATTTTCTGATTGAAATGATGC----- 158
Db 1623 eProGly-----CysGluThrThrIleGluIleSerLy 1634
QY 159 ----AGAGAAGGACTATCTATGATGCTGTG----- 186
Db 1634 sclyArgThrGlyLeuGlyLeu-SerIleValGlyGlySerAspThrLeuLeuGlyAlaI 1654
QY 187 -----CGATGTACACCCAG-----ACCATGGACGTGGCGGTGCTGCTG 226
Db 1654 leIleIleHisGluValTyrGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaG 1674
QY 227 GAGACTGAAGCTGTGTCATCAAT----- 249
Db 1674 lyAspGlnIleLeuGluValAsnGlyIleAspLeuArgLysAlaThrHisAspGluAlaI 1694
QY 250 -----GAACCCAGCTCTGCTCTGTTGATGCCATTCGCGCGCTGA 292
Db 1694 leAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyrArgAspGluA 1713
```

```
QY 293 TCCCACTGAAGCAGCAGGTGGTAATATGATCAGCTGACCGCCCGCGCTCCAGGAAGCTGA 352
Db 1713 laProTyrrLySgluGluGluValCysaspThrLeuThr-----
QY 353 AGGAGGTGCGTCTGACCGCTGCGACCGCTGCGCCAGCGCTGCGCTGAGTGTGGTGGTGCC 412
Db 1726 -----IleGluLeuGlnLysLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA 1744
QY 413 TGGAGTTGGCTGGGTCTTCATCTCCACCTCATCATCAAGGGCGGTGAGCAGCAGCG 472
Db 1744 rgAsn---AspThrGlyValPheValSerAspIleValLysGlyGlyIleAlaAspAlaA 1763
QY 473 TCGGG---CTCAGGTAGGAGCAGATCTCGCGATCAATGGATFATTCCTCTCCTCCT 529
Db 1763 spGlyArgLeuMetGlnGlyaspGlnIleLeuMetValasnGlyGluaspValargAsnA 1783
QY 530 GTACCATGAGGAGGTATCAACCTCATTCGA---ACCAAGAAACTGTGTCCATCAAG 586
Db 1783 laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGlyThrValThrLeuGluV 1803
QY 587 TCAGACACATCGCCTGATCCCGTGAAGCTCTCCTGATGAGCCCTCACTTGGCAGT 646
Db 1803 al-----GlyArgIleLysAlaGlySerSerThrSerGluSerLeu----- 1816
QY 647 ATGTGGATCAGTTTGTGCGGAATCTGGGGCGCTGCGAGCAGCCTGGGCTCC-----C 700
Db 1817 -----GluSerSerLysLysAsnAlaLeuAlaSerGluIleG 1830
QY 701 CTGGAATCGGAAACACAGAGAGAGAGTCTTCATCAGCCTGTGTAGCTCCCGAGGCC 760
Db 1830 lngLeuArgThrValGluMetLysLys-----GlyProThrAspSerL 1845
QY 761 TTGGCTGTCAGATTTCCACCGCC-----CCCATCCAGAGCCTGCATCTTTATCA 811
Db 1845 euGlyIleSerIleAlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleA 1865
QY 812 GCATGTGAACCTGCTCCCTGCTGCTGAGGTG---GGATGGAGATAGGGGACCCAGA 868
Db 1865 laMetHisProThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgI 1885
QY 869 TTGTCAAGTCAATGGCTGCTGCTCTCACTTCACTGATGATCACAAGAGGCTGTAATGTCC 928
Db 1885 leValThrIleCysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValasnLeuL 1905
QY 929 TGAATAATAGCCGACCTGACCATCTCCATTTAGCTGCGAGCTGCGCGGAGCTTTCAC 988
Db 1905 euLysasnAla---SerGlySerIleGluMetGlnValAlaGly-GlyaspValSer 1923
QY 989 TGACAGACGGGAGCGGCTGGCAGAGCGCGCAGCGTGTGAGTGCAGCGGAGGAGCTTC 1048
Db 1924 ValValThrGly----- 1927
QY 1049 TCATGCAGAGCGGCTGGCGATGGATCCCAACAGATCCTCCAGGAGCAGCAGGATGG 1108
Db 1927 ----- 1927
QY 1109 AGCGGCAAGGAGAGAAATTTGCCAGAGGAGCAGAGAGGAAATGAGAGATACCGGA 1168
Db 1927 ----- 1927
QY 1169 AGGAGATGGAACAGATTGTAGAGGAGAGAGAGATTAAAGAACTGGGAAGAAGACT 1228
Db 1927 ----- 1927
QY 1229 GGGCTCAAGGAAACAGCTACTCTTGCCTAAACCATCACTGCTGAGGTACACCCAGTAC 1288
Db 1928 -----HisGlnGlnGlu 1931
QY 1289 CCTTCGCAAGCAAGTATTTCCGGAATATGAGGAAGGCTTGAACCCCTACTCTATG 1348
Db 1932 ProAlaSerSerSerLeuSerPhe-----ThrGlyLeuThrSerSerSerIle 1947
QY 1349 TTCACC-----CCAGAGCAGATCATGGGAAGGATGTCCGGCTCTTACGCATC 1396
```

```
Db 1948 PheGlnAspAspLeuGlyProGlnCys-----LysSerIleThrLeuGluArg--- 1964
QY 1397 AAGAGGAGGATCCTTAGACCTGGCCCTGAAAGCGGTGGACTCCCCCATTTGGG--- 1453
Db 1965 ---GlyProaspGlyLeuGlyPheSerIleValGlyGlyThrGlySerProHisGlyAsp 1983
QY 1454 ---AAGGTGGTCTTCTGCTGATGAGCGGGAGCTGCTGAGCGGCATGTGGCAATT 1510
Db 1984 LeuProIleThrValLysThrValPheAlaLysGlyAlaAlaLaserGluAspGlyArgLeu 2003
QY 1511 GTAAAGGGAGCAGATCATGCGCAATCAAGCGCAAGATTGTGACAGACTACACCTGGCT 1570
Db 2004 LysArgGlyaspGlnIleIleAlaValasnGlyGlnSerLeuGluGlyValThrHisGlu 2023
QY 1571 GAGCTGACGCTGCCCTGCAGAAG 1594
Db 2024 GluAlaValAlaIleLeuLysArg 2031
RESULT 13
Q9NTP3 PRELIMINARY; PRT; 202 AA.
ID Q9NTP3;
AC Q9NTP3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DB BAI08L7.3 (Novel PDZ (DHR, GLCF) domain protein) (Fragment).
GN BAI08L7.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Heath P.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133215; CAB92801.1; -.
DR HSSP; P29476; 10AV.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
FT NON_TER 1 202
FT NON_TER 202 202
SQ SEQUENCE 202 AA; 21709 MW; 50B6507836F8845C CRC64;
Alignment Scores:
Pred. No.: 2, 22e-11 Length: 202
Score: 272.00 Matches: 69
Percent Similarity: 52.74% Conservative: 37
Best Local Similarity: 34.33% Mismatches: 83
Query Match: 6.82% Indels: 12
DB: 4 Gaps: 5
US-09-502-945-5 (1-2162) x Q9NTP3 (1-202)
QY 379 CCCGAAGGC---CTCGGCTGAGTGTGGTGGCTGGCTGGAGTTGGCTGGGCTCTTC 435
Db 6 ProAlaGlyArgLeuGlyPheSerValArgGlyGlySerGluHisGlyLeuGlyIlePhe 25
QY 436 ATCTCCACCTCATCAAGGGCGGTGAGCAGACAGCGTCCGGCTCCAGGTAGGAGCAG 495
Db 26 ValSerLysValGluGlySerSerAlaGluArgAlaGlyLeuCysValGlyAspLys 45
QY 496 ATCTCGCGGATCAATGGATATTCTCTCTCTTACCATGAGGAGGTCTATCAACCTC 555
Db 46 IleThrGluValasnGlyLeuSerLeuGluSerThrThrMetGlySerAlaValLysVal 65
QY 556 ATTCGAACCAAGAAACTGTGTCCATCAAGAGTGAGACATCGGCTGTATCCCTGAAA 615
Db 66 LeuThrSerSerArgLeuHisMetValArgMetGlyArgMetGlyArgValProGlyIle 85
QY 616 AGCTCTCTGTAGTGGCCCTCACTTGGCAGTATGTGGAT-----CAGTTGTG 663
```

```

Db 86 LysPheSerLysGluLysThrThrTrp-----ValAspValValAsnArgArgLeuVal 103
QY 664 TCGGAATCTGGGGCGTGGAGGCGAGCTGGGCTCCCTCGGAAATCGGAAACAAAGAG 723
Db 104 ValGluLys-----CysGlySerThrProSerAspThrSerSerGluAspGlyVal 120
QY 724 AAGAGGTCTTCATCAGCCTGGTAGGCTCCCGAGGC-----CTTGGCTGCAGCATTTCC 777
Db 121 ArgArgIleValHisLeuThrThrThrSerAspAspPheCysLeuGlyPheAsnIleArg 140
QY 778 AGCGCCCATCCAGAACCTGGCATCTTTATCAGCCATGTGAACCTCGGCTCCCTGTCT 837
Db 141 GlyGlyLysGluPheGlyLeuGlyIleThrValSerLysValAspHisGlyGlyLeuAla 160
QY 838 GCTAGGTGGGATTGGACATAGGAGCAGCATGTGCAAGTCAATGGCGTGCAGCTTCTCT 897
Db 161 GluGluAsnGlyIleLysValGlyAspGlnValLeuAlaAlaAsnGlyValArgPheAsp 180
QY 898 AACCTGGATCAAGAGGCTGTAATGTCTGTAATAATAGCCGAGCCTGACCATCTCC 957
Db 181 AspIleSerHisSerGlnAlaValGluValLeuLysGlyGlnThrHisIleMetLeuThr 200
QY 958 ATT 960
Db 201 Ile 201

RESULT 14
Q96C69 PRELIMINARY; PRT; 835 AA.
ID Q96C69
AC Q96C69;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 87.7 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014632; AAH14632.1;
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR PROSITE; PS0106; PDZ; 3.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 835 AA; 87669 MW; F59DF419D7AAF991 CRC64;

Alignment Scores:
Pred. No.: 4.33e-11 Length: 835
Score: 271.50 Matches: 178
Percent Similarity: 32.13% Conservative: 90
Best Local Similarity: 21.34% Mismatches: 252
Query Match: 6.81% Indels: 316
DB: 4 Gaps: 31

US-09-502-945-5 (1-2162) x Q96C69 (1-835)
QY 289 CTGATCCACTGAAGCACCAGGTGGATATGATCAGCTGACCCCGCGCTCCAGGAAG 348
Db 30 IleThrProLeuArgProGluAspAspThrSerPro-----ArgGluArgArg 45
QY 349 CTGAAGAGGTGGCTGTGGACCGCTCTCCACCCGAA----- 384
Db 46 GlyGlyGlyLeuArgLeuProLeuLeuProGluSerProGlyProLeuArgGlnArg 65
QY 385 -----GGCTCGCGCTGAGTGTGGTGGC----- 411
Db 66 HisValAlaCysLeuAlaArgSerGluArgGlyLeuGlyPheSerIleAlaGlyGlyLys 85

```

```

QY 412 -----CTGGAGTTTGGCTGGGGCTCTTCTATCTCCACCTCATCAAGGC 456
Db 86 GlySerThrProThrArgAlaGlyAspAlaGlyIlePheValSerArgIleAlaGluGly 105
QY 457 GGTGAGGAGCAGACGCTCGG-----CTCAGGTAGGAGCAGAGATCGTCCGGATCAATGGA 513
Db 106 GlyAlaAlaHisArgAlaGlyThrLeuGlnValGlyAspArgValLeuSerIleAsnGly 125
QY 514 TATTCCATCTCTCTCTACCCATGAGGAGGTCTCAACCTCATTT-----CGAACCAAGAAA 570
Db 126 ValAspValThrGluAlaArgHisAspHisAlaValSerLeuLeuThrAlaAlaSerPro 145
QY 571 ACTGTGTCCATCAAGTGCAGACATCGGCTGTATCCCGGTGAAGACTCTCTCT 624
Db 146 ThrIleAlaLeuLeuGluArgGluAlaGlyGlyProLeuProProSerProLeuPro 165
QY 625 -----GATGAGCCCTCACTTGGCAGTATGTGTCAGTGTGTCGGAATCTGGGGC 678
Db 166 HisSerSerProThrAlaAlaValAlaThrThrSerIleThrAlaThrProGly 185
QY 679 GTGCGAGGC----- 687
Db 186 ValProGlyLeuProSerLeuAlaProSerLeuLeuAlaAlaLeuGluGlyProTyr 205
QY 688 -----AGCTGGGCTCCCTCGAATCGGAAATCGGAAACACAGGAG 723
Db 206 ProValGluGluIleArgLeuProArgAlaGlyPro----- 218
QY 724 AAGAAGCTCTTCATCAGCTGTAGGCTCCCGAGGCTTGTGTCAGCATTTCCAGCGGC 783
Db 219 -----LeuGlyLeuSerIleValGly-----GlySerAspHisSerSerHis 232
QY 784 CCC-----ATCCAGAACCTGGCATCTTTATCAGCCATGTGAACCTGCTCTCTCT 837
Db 233 ProPheGlyValGlnGluProGlyValPheIleSerLysValLeuProArgGlyLeuAla 252
QY 838 GCTGAGGTGGATTGGAGATAGGAGGACCATGTGCAAGTCAATGGCTGCTGCTCTCT 897
Db 253 AlaArgSerGlyLeuArgValGlyAspArgIleLeuAlaValAsnGlyGlnAspValArg 272
QY 898 AACCTGGATCAAGAGGCTGTAATGTCTG----- 930
Db 273 AspAlaThrHisGlnGluAlaValSerAlaLeuLeuArgProCysLeuGluLeuSerLeu 292
QY 930 ----- 930
Db 293 LeuValArgArgAspProAlaProProGlyLeuArgGluLeuCysIleGlnLysAlaPro 312
QY 931 -----AAAAATACCCGC 942
Db 313 GlyGluArgLeuGlyIleSerIleArgGlyGlyAlaArgGlyHisAlaGlyAsnProArg 332
QY 943 AGCCTGACC-----ATCTCCATTGTAGCTGCAGTGCCTGCGCG 978
Db 333 AspProThrAspGluGlyIlePheIleSerLysValSerProThrGlyAlaAlaGlyArg 352
QY 979 GAGCTGTTCATGACACCGGAGCGGCTGGCAGAGCGCGGCGGCGTGTGAGTGTG----- 1032
Db 353 AspGlyArgLeuArgValGlyLeuArgLeuLeuGluValAsnGlnGlnSerLeuLeuGly 372
QY 1033 -----CAGCGGAGGAGCTTCTCATGCA----- 1055
Db 373 LeuThrHisGlyGluAlaValGlnLeuLeuLeuLeuLeuValGlyAspThrLeuThrValLeu 392
QY 1056 -----GAAGCGCTGCGGATGGAGTCCACAGATCCTCCAGGA----- 1094
Db 393 ValCysAspGlyPheGluAlaSerThrAspAlaAlaLeuGluValSerProGlyValIle 412
QY 1095 -----GCAGCAGGAGATGGA-----CGCGCAAG 1118
Db 413 AlaAsnProPheAlaAlaGlyIleGlyHisArgAsnSerLeuLeuSerIleSerIle 432

```


Db 835 GlyAlaAlaHisArgAlaGlyThrLeuGlnValGlyAspArgValLeuSerIleAsnGly 854
QY 514 TATTCATCTCTCTCTACCCATGAGGAGGTCAACACCTCATTT---CGAACCAAGAAA 570
Db 855 ValAspValThrGluAlaArgHisAspHisAlaValSerLeuLeuThrAlaAlaSerPro 874
QY 571 ACTGTGTCCTCAAAAGTGAACACATCGGCCTGTATCCCGCTGAAAGCTCTCCT 624
Db 875 ThrIleAlaLeuLeuGluAlaGluAlaGlyGlyProLeuProProSerProLeuPro 894
QY 625 -----GATGAGCCCTCACITGGCAGTATGTGGATCAGTTGTGTCGGAATCTGGGGC 678
Db 895 HisSerSerProThrAlaAlaValAlaThrThrSerIleThrAlaThrProGly 914
QY 679 GTGCGAGGC----- 687
Db 915 ValProGlyLeuProSerLeuAlaProSerLeuLeuAlaAlaAlaLeuGluGlyProTyr 934
QY 688 -----AGCCTGGGCTCCCTGGAAATCGGGAACCAAGAG 723
Db 935 ProValGluGluIleArgLeuProArgAlaGlyGlyPro----- 947
QY 724 AAGAGGCTTCATCAGCCTGGTAGGCTCCGAGGCTTGCTGCAGCATTTCCAGCGGC 783
Db 948 -----LeuGlyLeuSerIleValGly-----GlySerAspHisSerHis 961
QY 784 CCC-----ATCCAGAGCCCTGATCTTTATCAGCCATGTGAACCTGGCTCTCTCT 837
Db 962 ProPheGlyValGlnGluProGlyValPheIleSerLysValLeuProArgGlyLeuAla 981
QY 838 GCTGAGGTGGATGAGGACACAGATTTCGAAGTCAATGGCTGACTTCTCT 897
Db 982 AlaArgSerGlyLeuArgValGlyAspArgIleLeuAlaValAsnGlyGlnAspValArg 1001
QY 898 AACCTGATCACAGGAGGCTGTAATGTGCTG----- 930
Db 1002 AspAlaThrHisGlnGluAlaValSerAlaLeuLeuArgProCysLeuGluLeuSerLeu 1021
QY 930 ----- 930
Db 1022 LeuValArgArgAspProAlaProProGlyLeuArgGluLeuCysIleGlnLysAlaPro 1041
QY 931 -----AAAAATAGCCGC 942
Db 1042 GlyGluArgLeuGlyIleSerIleArgGlyGlyAlaArgGlyHisAlaGlyAsnProArg 1061
QY 943 AGCCTGACC-----ATCTCCATTGTAGCTGAGCTGCGCGG 978
Db 1062 AspProThrAspGluGlyIlePheIleSerLysValSerProThrGlyAlaAlaGlyArg 1081
QY 979 GAGCTGTTCATACAGACCGGAGCGGCTGCGAGAGCGCGGCGCTGAGCTG----- 1032
Db 1082 AspGlyArgLeuArgValGlyLeuArgLeuLeuGluValAsnGlnGlnSerLeuLeuGly 1101
QY 1033 -----CAGCGGCAGGAGCTTCTCATGCA----- 1055
Db 1102 LeuThrHisGlyGluAlaValGlnLeuLeuArgSerValGlyAspThrLeuThrValLeu 1121
QY 1056 -----GAAGCGGTGGCGATGGAGTCCAAACAGATCCTCCAGGA----- 1094
Db 1122 ValCysAspGlyPheGluAlaSerThrAspAlaAlaLeuGluValSerProGlyValIle 1141
QY 1095 -----GCAGCAGGATGGA-----GCGCAAG 1118
Db 1142 AlaAsnProPheAlaAlaGlyIleGlyHisArgAsnSerLeuGluSerIleSerIle 1161
QY 1119 GAGAAAGAAATTCGCCAGAGGACGACAGAGGAAATGAGATACCGAGGAGGATGGA 1178
Db 1162 AspArgGluLeuSerProGluGlyProGlyLysGluLysGluLeuPro-----Gly 1178
QY 1179 ACAGATTGTAGAGGAGAGAGGATTTAAGACCAATGGGAAGA----- 1223
Db 1179 GlnThrLeuHisTrpGlyProGluAlaThrGluAlaAlaGlyArgGlyLeuGlnProLeu 1198

QY 1224 AGACTGGGCTCAAGGAACAGCTACTCTTCCTGCCTAAAAACCATCACTGCTGAGGTACACCC 1283
Db 1199 LysLeuAspTyrArgAlaLeuAlaVal-----Pro 1209
QY 1284 AGTACCCTTCGCAAGCAAGTGTATTCGGAATATGAGGAAGGCTTTCACCCCTACT 1343
Db 1210 SerAlaGlySerValGlnArgValProSerGlyAlaAlaGlyLysMetAlaGluSer 1229
QY 1344 CTATGTTCAACCCAGACAGCATCATGGGAAGGATGTCCGGCTCTACGCTCAAGAAGG 1403
Db 1230 ProCysSerProSerGlyGln-GlnProProSerProSerProAspGluLeuProAl 1249
QY 1404 AGGATCCTTAGA----- 1416
Db 1249 aAsnValLysGlnAlaTyrArgAlaPheAlaAlaValProThrSerHisProProGluAs 1269
QY 1417 -----CCTGGCCCTGGNAGG----- 1431
Db 1269 pAlaProAlaGlnProProThrProGlyProAlaAlaSerProGluGlnLeuSerPheAr 1289
QY 1432 -----CGTGTGACTCCCCCATTTGGGAAGT-----GG 1460
Db 1289 gGluArgGlnLysTyrPheGluLeuGluValArgValProGlnAlaGluGlyProProLy 1309
QY 1461 TCGTTTCTGCTGTATGACGGGGAGCTCTGAGCGGCATGTGTCATTGTGAAAGGG 1520
Db 1309 sArgValSerLeuVal----- 1315
QY 1521 ACNAGATCATGGCANTCAACGGCAAGNTTGTGACAGATACACCCCTGGCTGAGGCTGACG 1580
Db 1315 yAlaAspAspLeuArgLysMetGlnGlu----- 1324
QY 1581 CTGCGCTGCAGAAGCGCTGGAATCAGGGGGGACTGGATCGACCTTGTGTTGCCCTCT 1640
Db 1325 -----GluGluAlaArgLysLeuGlnGln----- 1332
QY 1641 GCCCCCAAGGAGTATGACGATGACCTGACCTTCTGTGAAGTCCAAAGGGGAAACC 1700
Db 1333 -----LysArgAlaGlnMetLeuArgGluAlaAlaGlyAl 1346
QY 1701 AAATTCACGCGTTAGGAACACAGTGAGCTCCGGCCCCACCTCGTGAACACAAACCTCGGA 1760
Db 1346 aGluAlaArgLeuAlaLeuAspGlyGluThrLeuGlyGluGluGlnGluAspGluG 1366
QY 1761 CCAGCCTTCAGAGAGCCACATGACACACAGATGGCATCTCTGGGACCT----- 1812
Db 1366 nProProTrpAlaSerProSerProThrSerArgGlnSerProAlaSerProProProLe 1386
QY 1813 -----GAATCTATCACCCAGGAATCTCAAC 1838
Db 1386 uGlyGlyGlyAlaProValArgThrAlaLysAlaGluArgArgHisGlnGluArgLeuAr 1406
QY 1839 TCCTTTTGGCCCTGAACAGGCGCCAGATAGGAACAGCTCGGGCCACTTTT-TTGAAGCG 1897
Db 1406 gValGlnSerProGluProAlaProAlaProGluAlaLeuSerProAlaGluLeuArgAl 1426
QY 1898 CAATGTGGGAAAGGAGCAGCCAGCCCTTTGGGAGAAGATCTCAAGGATCCAGACTCT 1957
Db 1426 a-----LeuGluAlaGluLysArgAlaLeuTrp---ArgAlaAlaArgMetLysSerLe 1443
QY 1958 CATTCCTTTCTCTGTCGCCAGTGAATTTGCTCTCTCCAGCTTTGG-----GGGAC 2008
Db 1443 uGluGlnAspAlaLeuArgAlaGlnMetValLeuSerArgSerGlnGluGlyArgGlyTh 1463
QY 2009 T-----CCTTCCTTGAACCCCTAATAAGAC 2032
Db 1463 rArgGlyProLeuGluArgLeuAlaGluAlaProSerProAlaProThrProSerProTh 1483
QY 2033 CCCACTGGAGTCTCTC----- 2048
Db 1483 rProValGluAspLeuGlyProGlnThrSerThrSerProGlyArgLeuSerProAspPh 1503

Tue Mar 25 09:28:42 2003

Qy 2049 -----TCTCTCCATCCCTCTCTGCCC 2072
||||| ||||| |||
Db 1503 eAlaGluLeuArgSerGluProSerProSerPro 1516

Search completed: March 21, 2003, 12:59:57
Job time : 103.143 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 11:21:37 ; Search time 27.1901 Seconds
(without alignments)
8801.668 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346

Sequence: 1 ggaattcctctgttcgaagt.....aaacaaaagttaaaattt 2885

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlpl
-Q/cgn2.1/USPTO_spool/US09502945/runat_14032003_101058_19100/app_query.fasta_1.10979
-DB=SwissProt40 -OPMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09502945@cgn.1.113@runat_14032003_101058_19100 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	4793	89.7	1122	1 HDA5_HUMAN	Q9uql6 homo sapien
2	4565.5	85.4	1113	1 HDA5_MOUSE	Q92zv6 mus musculus
3	2732.5	51.1	1080	1 HDA4_CHICK	P83038 gallus gall
4	2691	50.3	1084	1 HDA4_HUMAN	P56524 homo sapien
5	2579	48.2	1011	1 HDA9_HUMAN	Q9ukv0 homo sapien
6	950.5	17.8	588	1 HDA9_MOUSE	Q99n13 mus musculus
7	771.5	14.4	1215	1 HDA6_HUMAN	Q9ubn7 homo sapien
8	771	14.4	1149	1 HDA6_MOUSE	Q92zv5 mus musculus
9	705.5	13.2	687	1 HDA1_SCHPO	P56523 schizosacch
10	672	12.6	706	1 HDA1_YEAST	P53973 saccharomyc
11	651.5	12.2	798	1 YLFN_CAEEL	Q20296 caenorhabdi
12	490.5	9.2	359	1 Y130_ARCFU	O30107 archaeglob
13	385	7.2	331	1 Y894_METHH	O27262 methanobact
14	358.5	6.7	304	1 Y245_SYNY3	P72702 synechocyst
15	348	6.5	343	1 Y535_METJA	O57955 methanococ
16	345	6.5	385	1 ACUC_STAXY	O56195 staphylococ
17	331.5	6.2	387	1 ACUC_BACSU	P39067 bacillus su
18	316.5	5.9	480	1 HDA1_CHICK	P56517 gallus gall

19	312	5.8	461	1 HDA1_CAEEL	017695 caenorhabdi
20	309	5.8	488	1 HDA2_MOUSE	P70288 mus musculus
21	306.5	5.7	482	1 HDA1_HUMAN	Q13547 homo sapien
22	306.5	5.7	482	1 HDA1_MOUSE	Q09106 mus musculus
23	304	5.7	488	1 HDA2_HUMAN	Q92769 homo sapien
24	303.5	5.7	433	1 RPD3_YEAST	P32561 saccharomyc
25	303	5.7	480	1 HD11_XENLA	Q91695 xenopus lae
26	301	5.6	452	1 HOS2_YEAST	P53096 saccharomyc
27	297.5	5.6	434	1 PHD1_SCHPO	O13298 schizosacch
28	297	5.6	480	1 HD12_XENLA	O42227 xenopus lae
29	289	5.4	576	1 HDA1_STRPU	P56518 strongyloe
30	288	5.4	513	1 HDAC_MAI2E	P56521 zea mays
31	285	5.3	520	1 HDAC_DROME	Q94517 drosophila
32	280	5.2	488	1 HDA2_CHICK	P56519 gallus gall
33	279.5	5.2	660	1 YHL1_EBV	P03181 epstein-bar
34	268	5.0	507	1 HDA2_CAEEL	Q09440 caenorhabdi
35	265.5	5.0	501	1 HDAC_ARATH	O22446 arabidopsis
36	261	4.9	424	1 HDA3_MOUSE	O88895 mus musculus
37	260	4.9	428	1 HDA3_CHICK	P56520 gallus gall
38	260	4.9	428	1 HDA3_HUMAN	O15379 homo sapien
39	255.5	4.8	660	1 YHL1_EBV	P03181 epstein-bar
40	254.5	4.7	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
41	253.5	4.7	2715	1 TRX2_HUMAN	O9um06 homo sapien
42	251	4.7	341	1 APHA_MYCRA	O48935 mycoplasma r
43	242	4.5	2161	1 SHK1_HUMAN	O9y566 homo sapien
44	239	4.5	1733	1 VNVA_PPVKA	P33485 pseudorabie
45	234	4.4	1783	1 RAA3_CHLRE	Q9fec4 chlamydomon

ALIGNMENTS

RESULT 1
HDA5_HUMAN STANDARD: PRT: 1122 AA.
AC Q9UQL6; O60528; O60340;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone deacetylase 5 (HDS) (Antigen NY-CO-9).
GN HDAC5 OR KIA0600.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-9238449; PubMed-10220385;
RA Grozinger C.M., Hassig C.A., Schreiber S.L.;
RT "Three proteins define a class of human histone deacetylases related to yeast Hda1p.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:4868-4873(1999).
RN [2]
RP SEQUENCE OF 189-1085 FROM N.A.
RC TISSUE-Colorectal carcinoma;
RA Scanlan M.J., Chen Y., Williams B., Gure A.O., Stockert E.,
RT "Characterization of human colon cancer antigens recognized by autologous antibodies.";
RN [3]
RP SEQUENCE OF 407-1122 FROM N.A.
RC TISSUE-Brain;
RX MEDLINE-98290545; PubMed-9628581;
RA Nagase T., Ohkawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RT "Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";
RL DNA Res. 5:31-38(1998).
CC -!- FUNCTION: FOR THE DEACETYLATION OF LYSINE RESIDUES ON THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4). HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL

CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS.
 CC REPRESSORS TRANSCRIPTION (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
 CC FAMILY, HD SUBFAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF132608; AAD29047.1; -;
 CC EMBL: AF039691; AAC18040.1; -;
 CC EMBL: AB011172; BAA25526.1; -;
 CC Genbank: HGNC:14068; HDAC5..
 CC MIN: 605315; -;
 CC InterPro: IPR000286; His_deacetylase.
 CC Pfam: PF00850; Hist_deacetyl; 1.
 CC PRINTS: PR01270; HDASUPER.
 CC K W Hydrolyase; Nuclear protein; Transcription regulation; Repressor.
 FT DOMAIN 47 52 POLY-GLY.
 FT DOMAIN 85 92 POLY-GLN.
 FT DOMAIN 596 599 POLY-GLU.
 FT DOMAIN 1099 1104 POLY-ALA.
 FT DOMAIN 684 1028 HISTONE DEACETYLASE.
 FT CONFLICT 671 671 S -> N (IN REF. 2).
 FT CONFLICT 684 684 G -> S (IN REF. 2).
 FT CONFLICT 1026 1026 E -> K (IN REF. 2).
 FT CONFLICT 1074 1074 E -> G (IN REF. 2).
 FT CONFLICT 1085 1085 T -> M (IN REF. 2).
 FT CONFLICT 1122 AA; 121991 MW; CF4BBB8B9A288FEC CRC64;
 SQ SEQUENCE

Alignment Scores:

Pred. No.: 3 43e-193 Length: 1122
 Score: 4793.00 Matches: 927
 Percent Similarity: 99.57% Conservative: 3
 Best Local Similarity: 99.25% Mismatches: 4
 Query Match: 89.66% Indels: 1
 DB: 1 Gaps: 0

US-09-502-945-2 (1-2885) x HDA5_HUMAN (1-1122)

QY 2 GAATTCCTCTTCGAGTCAAGAGGCCACACGAGGGCGCTCAACATTCCTCCCA 61
 Db 189 GluPheLeuSerLysSerLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 208
 QY 62 CAGCACCCAAATGCTGGGAGCCACCATGCTCTTTGGACGAGAGTCCCTCCCGAG 121
 Db 209 GlnHisProLysCysTrpGlyAlaHisAlaSerLeuAspGlnSerSerProGln 228
 QY 122 AGCGGCCCCCTGGGAGCGCTCCCTACAAACTGCCTTTGCTGGGCGCTTACGACAGT 181
 Db 229 SerGlyProProGlyThrProProSerTyLysLeuProLeuProGlyProTyLysSer 248
 QY 182 CGAGAGACTTCCCTCCGCAAAACAGCCTCTGACCCAACTGAAAGTGGCTTCAAGG 241
 Db 249 ArgAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg 268
 QY 242 CTAACAACAAGGTGCTGAGCGGAGAACAGTCCCTCTCCGTCGCAAGGATGGACT 301
 Db 269 LeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgLysAspGlyThr 288
 QY 302 GTTATTAGCACCTTTAAGAAAGAGAGCTGTGATCACAGGTGCCGGGCTGGGGCGTG 361
 Db 289 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyAlaGlyProGlyAlaSer 308
 QY 362 TCCGTGTGAACAGCCAGCCCGGCTCCGGCCCCAGCTCTCCCAACAGCTCCACAGCACC 421
 Db 309 SerValCysAsnSerAlaProGlySerGlyProSerSerProAsnSerSerHisSerThr 328

QY 422 ATCGCTGAGAATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCAG 481
 Db 329 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGln 348
 QY 482 CACCGAGCCCTCCCTCTGGACAGCTCCCAACAGTTCAGCTCTACAGCTCTCTTCT 541
 Db 349 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTyThrSerProSer 368
 QY 542 CTGCCCCAACATCTCCCTAGGCTGCAGGCCACCGTCTCCTCACCACCTCACCCTCCT 601
 Db 369 LeuProAsnIleSerLeuGlyLeuGlnAlaThrValThrValThrAsnSerHisLeuThr 388
 QY 602 GCCTCCCGGAAGTGTGACACAGCAGGAGCGGAGGAGCGGAGGAGGAGGAGGAGGAG 661
 Db 389 AlaSerProLysLeuSerThrGlnGlnAlaGluArgGlnAlaLeuGlnSerLeuArg 408
 QY 662 CAGGTGGCAGCTGACCGGCAAGTTCATGAGCAGCATCTCTATTCTGCTGGCTGCTG 721
 Db 409 GlnGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCysLeuLeu 428
 QY 722 GCGTGGCAGTGGAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 781
 Db 429 GlyValAlaLeuGluGlyAspGlySerProHisGlyHisAlaSerLeuLeuGlnHisVal 448
 QY 782 CTGTTGCTGGAGGAGCGGCGGAGCAGCAGCACCTCTATTGCTGTGCCACTCCACGGG 841
 Db 449 LeuLeuLeuGluGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 468
 QY 842 TCCCCACTAGTACGGGTGAACGTGTGGCCACGACGATGCGGAGGAGGAGGAGGAGG 901
 Db 469 SerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLysLeuPro 488
 QY 902 CGGCATCGGCCCTGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 961
 Db 489 ArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGlnAlaLeu 508
 QY 962 CAGCAGCTGTGTCATGCAACACAGCAGCAGCAGGAGTCTTCTGGAGAGCAGACAGCAG 1021
 Db 509 GlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnGlnGln 528
 QY 1022 CTACAGCTGGGCAAGATCTTCACCAAGACAGGGGAGGAGTCCCGAGGAGCGGAGGAG 1081
 Db 529 LeuGlnLeuGlyLysIleLeuThrLysThrGlyLeuLeuProArgGlnProThrHis 548
 QY 1082 CTGAGGAGCAGAGGAGGAGGAGTCCAGCAGCAGCAGGAGTCTTCTGGGAGGAGGAGCC 1141
 Db 549 ProGluGluThrGluGluGluLeuThrGluGlnGlnGluValLeuLeuGlyGlyAla 568
 QY 1142 CTGACCATGCCCCGGGAGGCTCCACAGAGAGTGGAGGAGCAGCAGAGGAGGAGGAG 1201
 Db 569 LeuThrMetProArgGluGlySerThrGluSerGluSerThrGlnGluAspLeuGlu 588
 QY 1202 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1261
 Db 589 GluAspGluGluGluAspGlyGluGluGluAspGlyGluGluGluGluGluGluGlu 608
 QY 1262 GCGCAGAGTGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1321
 Db 609 GlyGluSerGlyAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 628
 QY 1322 CTGTTCTCAGATGCCCAACCGCTGCAACCTTTGAGGTGTACCAAGGCGGCGGCTCAG 1381
 Db 629 LeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyGlnAlaProLeuSerLeu 648
 QY 1382 GCCACTGTGCCCCACCAAGCCCTGGGCGGTACCAATCTCCCTCCCTGCTGCGGCGG 1441
 Db 649 AlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaProGlyGly 668
 QY 1442 ATGAAGAACCCCCCAGACCAACCGCTCAAGCAGCTCTCACCACAGTGTGGTCTAGCAG 1501
 Db 669 MetLysSerProProAspGlnProValLysHisLeuPheThrThrGlyValValTyAsp 688
 QY 1502 ACGTTTCATGCTAAAGCACCGAGTGCATGTGCGGGAACACACAGTGCACCTTGACATGCT 1561

Db	689	ThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGluHisAla	708
Qy	1562	GGCGGATCCAGACATCTGGTCCCGGCTGCAGGAGACAGGCGCTGTTAGCAAGTGCAG	1621
Db	709	GlyArgIleGlnSerIleTrpSerArgLeuGlnGluThrGlyLeuLeuSerLysCysGlu	728
Qy	1622	CGGATCCGAGGTCCCAAGCCAGCTAGATCAGATCCAGACAGTGCATCTCTGAATACCAC	1681
Db	729	ArgIleArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGluTrpHis	748
Qy	1682	ACCTGCTCTATGGACCACTGCCCTCAACGGCAGAACTAGACAGCAAGAAGTTGCTC	1741
Db	749	ThrLeuLeuTrpGlyThrSerProLeuAsnArgGlnLysLeuAspSerLysLysLeuLeu	768
Qy	1742	GGTCCCATCAGCCAGAAGATGTATGTGTGCTGCCCTGTGGGGCATCGGGGTGCACAGT	1801
Db	769	GlyProIleSerGlnLysMetTyrAlaValLeuProCysGlyGlyIleGlyValAspSer	788
Qy	1802	GACACCGTGTGGAATGAGTGCACCTCCACGAGTGTGCTGCCGATGGCAGTGGGTGCCTG	1861
Db	789	AspThrValIleTrpAsnGluMetHisSerSerSerAlaValArgMetAlaValGlyCysLeu	808
Qy	1862	CTGAGCTGGCCCTCAAGGTGGCTGCAGAGAGAGCTCAAGAATGGATTTGCCATCATCCGG	1921
Db	809	LeuGluLeuAlaPheLysValAlaAlaGlyGluLeuLysAsnGlyPheAlaIleLeuArg	828
Qy	1922	CCCCAGAGACACGCGGAGGAATCCACAGCCATGGGATTTGCTTCTTCAACTCTGTA	1981
Db	829	ProProGlyHisAlaGluSerThrAlaMetGlyPheCysPhePheAsnSerVal	848
Qy	1982	GCCATCACCCCAAACTCTACAGCAGAAAGTTGAAGTGGCGGCAAGTCCCTCATCGTGGAC	2041
Db	849	AlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeuIleValAsp	868
Qy	2042	TGGGACATTCACCATGGCAATGGCACCCAGCAGGCGTGTCTACAATGACCCCTGTGCTC	2101
Db	869	TrpAspIleHisGlyAsnGlyThrGlnGlnAlaPheTrpAsnAspProSerValLeu	888
Qy	2102	TACATCTCTCGATCGCTATGACAAACGGGAACCTCTTTCCAGGCTCTGGGCGCTCGTAA	2161
Db	889	TyrIleSerLeuHisArgTyrAspAsnGlnPhePheProGlySerGlyAlaProGlu	908
Qy	2162	GAGTTGGTGGAGACACGCGTGGGTACATGTGACGTGGCATGGACAGGAGTGTG	2221
Db	909	GluValGlyGlyGlyProGlyValGlyTyrAsnValAsnValAlaIleTrpThrGlyGlyVal	928
Qy	2222	GACCCCCCATTTGAGACGTGGAGTACCTTACAGCGCTTCAGCAGACGTGTGATCCCCATT	2281
Db	929	AspProIleGlyAspValGluTyrLeuThrAlaPheArgThrValValMetProIle	948
Qy	2282	GGCCACGAGTTCTACCTGATGGTGCCTAGTCTCCGCGGGTTGTATGCTGTGTGAAGGA	2341
Db	949	AlaHisGluPheSerProAspValValLeuValSerAlaGlyPheAspAlaValGluGly	968
Qy	2342	CATCTGTCCTCTGTGGTGGCTACTCTGTACCGCCAGATGTTTGGCCACTTGACCAGG	2401
Db	969	HisLeuSerProLeuGlyGlyTyrSerValThrAlaArgCysPheGlyHisLeuThrArg	988
Qy	2402	CAGCTGATGACCTTGCAGGGGCGGGTGGTGTGGCCCTGGAGGAGGCCATGACTTG	2461
Db	989	GlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGluGlyGlyHisAspLeu	1008
Qy	2462	ACGGCCATCTGTATGCCCTCTGAAGCTTGTGCTCGGCTGTGCTGCTCAGTGTAAAGCTCCAG	2521
Db	1009	ThrAlaIleCysAspAlaSerGluAlaCysValSerAlaLeuLeuSerValGluLeuGln	1028
Qy	2522	CCCTTGGATGAGCAGCTCTTTCAGCAAAAAGCCCAATCAACGACGTGGCCACGCTAGAG	2581
Db	1029	ProLeuAspGluAlaValLeuGlnGlnLysProAsnIleAsnAlaValAlaThrLeuGlu	1048
Qy	2582	AAAGTCATGAGATCCAGAGCAACACTGGAGCTGTGTGCAGAAAGTTCGGCGCTGGTCTG	2641

Db 1049 LysValIleGluIleInSerLysHisThrSerCysValInGlnLysPheAlaAlaGlyLeu 1068

QY 2642 GCCCGGTCCCTCGAGGGGCCCAACAGCTGAGACCGAGAGCCGAAAT-CTGAACGCC 2700

Db 1069 GlyArgSerLeuArgGluAlaGlnAlaGlyGluThrGluGluAlaGluThrValSerAla 1088

QY 2701 ATGGCCTTGCCTTTGTGGTGGGGCCCAACAGCGCCCAAGCTGCGGCAGCCGGGAACACACAG 2760

Db 1089 MetaAlaLeuLeuSerValGlyAlaGluGlnAlaGlnAlaAlaAlaArgGluHisSer 1108

QY 2761 CCCAGCGCGGACAGAGAGCCCATGAGCAGAGAGCCTGCCTCG 2802

Db 1109 ProArgProAlaGluGluProMetGluGlnGluProAlaLeu 1122

RESULT 2

HDAS_MOUSE

ID HDAS_MOUSE STANDARD; PRT: 1113 AA.

AC Q922V6; Q9JL73;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Histone deacetylase 5 (HD5) (Histone deacetylase mHDA1).

GN HDAC5

OS Mus musculus (Mouse).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID:10090;

FN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Fetal;

RA MEDLINE=99107904; PubMed=9891014;

RA Verdel A., Khochbin S.;

RT "Identification of a new family of higher eukaryotic histone

RT deacetylases. Coordinate expression of differentiation-dependent

RT chromatin modifiers.";

RT J. Biol. Chem. 274:2440-2445(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RA MEDLINE=20107033; PubMed=10640276;

RA Kao H.Y., Downes M., Ordentlich P., Evans R.M.;

RT "Isolation of a novel histone deacetylase reveals that class I and

RT class II deacetylases promote SMRT-mediated repression.";

RL Genes Dev. 14:55-66(2000).

CC -1- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON

CC THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).

CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL

CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS.

CC REPRESSSES TRANSCRIPTION.

CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA

CC FAMILY. HD SUBFAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC ENBL; AF006602; AAD09834.2; -.

DR EMBL; AF207748; AAF31418.1; -.

DR MGD; MGI:1333784; Hdac5.

DR InterPro; IPR000286; His_deacetylase.

DR Pfam; PF00850; Hist_deacetyl; 1.

DR PRINTS; PR01270; HDASUPER.

KW Hydrolase; Nuclear protein; Transcription regulation; Repressor.

FT DOMAIN 47 52 POLY-GLY.

FT DOMAIN 85 92 POLY-GLN.

FT DOMAIN 577 588 POLY-GLU.

FT DOMAIN 675 1019 HISTONE DEACETYLASE.

FT CONFICT 7 17 S -> SA (TN REF 2).

FT CONFLICT 18 18 G -> E (IN REF. 2).
SQ SEQUENCE 1113 AA; 120942 MW; 63071AF45B8715A CRC64;

Alignment Scores:

Pred. No.: 1.03e-183 Length: 1113
Score: 4565.50 Matches: 888
Percent Similarity: 96.70% Conservative: 19
Best Local Similarity: 94.67% Mismatches: 25
Query Match: 85.40% Indels: 7
DB: 1 Gaps: 2

US-09-502-945-2 (1-2885) x HDA5_MOUSE (1-1113)

QY 2 GAATTCCTCTTCGAGTCGAAGAGCCACACAGCGGCTCAACCATTCCTCCCA 61
DB 180 GluPheLeuLeuSerLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 199
QY 62 CAGCACCCCAATGCTGGGGAGCCACCATTGCTTTGGACAGAGTCCCTCCCGAG 121
DB 200 GlnHisProLysCysTrpGlyAlaHisAlaSerLeuAspGlnSerSerProGln 219
QY 122 AGCGGCCCCCGGGAGCGCTCCCTCCACAACTGCCCTTGCCTGGGCGCTACGACGT 181
DB 220 SerGlyProProGlyThrProProSerTyrLysLeuProLeuGlyProTyrAspSer 239
QY 182 CGAGACGACTTCCCTCCGCAAAACAGCCTCTGAACCCAACTGAAAGTGGCTCAAGG 241
DB 240 ArgAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg 259
QY 242 CTAAACAGAGGTGCTGAGCGGAGAGCAGTCCCTCCCTCGTCGCGCAAGGATGGACT 301
DB 260 LeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArgLysAspGlyThr 279
QY 302 GTTATTAGCACCTTTAAGAGAGAGCTGTGAGATCACAGGTGCCGGCGCTGGCGTCG 361
DB 280 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyThrGlyProGlyValSer 299
QY 362 TCCTGTGTAAACAGCGACCCCGGCTCCGGCCCGAGCTCTCCAAACAGCTCCACAGCAC 421
DB 300 SerValCysAsnSerAlaProGlySerGlyProSerSerProAsnSerSerHisSerThr 319
QY 422 ATGCTGTAGAATGGCTTTACTGGCTCAGTCCCGCAACATCCCGACTGAGTGTCCCTCAG 481
DB 320 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetIleProGln 339
QY 482 CACCGAGCCCTCCCTCTGACAGCTCCCGCCACCACTGCTAGCTCTACAGCTCTCCTCT 541
DB 340 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTyrThrSerProSer 359
QY 542 CTGCCCAACATCTCCCTAGGGCTGACGGCCACCGTCACTGTCAACCACTCACACCTCACT 601
DB 360 LeuProAsnIleSerLeuGlyLeuGlnAlaThrValThrValThrAsnSerHisLeuThr 379
QY 602 GCTTCCCGGAAGTGTGACACAGCAGGAGGCGGAGAGGCGCTCCAGTCCCTCGCGG 661
DB 380 AlaSerProLysLeuSerThrGlnGlnGluAlaGluArgGlnAlaLeuGlnSerLeuArg 399
QY 662 CAGGTGGCAGCTGACCGCAAGTTTCATGACGACATCCTCTATTCTGCTGCTCGCTG 721
DB 400 GlnGlyThrLeuThrGlyLysPheMetCysThrSerSerIleProGlyCysLeuLeu 419
QY 722 GGGTGGCACTGGAGGGCGAGCGGCCCGCCACGGGCATGCCCTCCCTGTCGACGATGTG 781
DB 420 GlyValAlaLeuGluGlyAspThrSerProHisGlyHisAlaSerLeuLeuHisVal 439
QY 782 CTGTTGCTGGAGGCGCGGACGAGACACCTTCATTTGCTGTGCTGCTCCAGCGGCGAG 841
DB 440 -CysSerTrpThr- -GlyArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 458
QY 842 TCCCGCACTAGTACGGGTGAACGTGTGGCCACGAGCATGCGGACGTAGCAAGCTCCCG 901
DB 459 SerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLysLeuPro 478

QY 902 CGGCATCGGCCCTGAGCGCAGCTCAGTCTCCTACCGCTGCCGAGAGTCCCGGCGCTG 961
DB 479 ArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGlnAlaLeu 498
QY 962 CAGCAGCTGGTCATGCAACAACACACACAGCTTCCTGGAGAGCAGACGACGACGAG 1021
DB 499 GlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnLysGlnGln 518
QY 1022 CTACAGCTGGGCAAGATCTCTACCAAGACAGGGAGCTGCCAGCAGCGCCACCCAC 1081
DB 519 MetGlnLeuGlyLysIleLeuThrLysThrGlyGluLeuSerArgGlnProThrHis 538
QY 1082 CCTGAGGAGACAGAGGAGCTGACGAGCAGCAGAGGTCTTCTGGGGGAGGAGCC 1141
DB 539 ProGluGluThrGluGluLeuThrGluGlnGlnAlaLeuLeuGlyGluGlyAla 558
QY 1142 CTGACCATGCCCCGGGAGGCTCCACAGAGTCCAGAGCAGCAGAGAGACTGGAGGAG 1201
DB 559 LeuThrIleProArgGluGlySerThrGluSerGluSerThrGlnGluAspLeuGluGlu 578
QY 1202 GAGGACGAGGAGGAGGATGGGAGGAGGAGGATTCATCCAGGTTAAAGCAGGAGGAG 1261
DB 579 GluGluGluGluGluGlu- -GluGluGluLysPheGlnValLysAspGluAsp 597
QY 1262 GCGGAGAGTGTGCTGAGGAGGCGGCTGAGGAGGAGCTGGTGTGGATCAAAAAA 1321
DB 598 GlyGluSerGlyProAspGluGlyProAspLeuGluGluSerSerAlaGlyTyrLysLys 617
QY 1322 CTGTTCTCAGATGCCCAACCGCTGCAACCTTTCAGGAGTACCAAGCCCTCAGCGCTG 1381
DB 618 LeuPheAlaAspAlaGlnGlnLeuGlnProLeuGlnValTyrGlnAlaProLeuSerLeu 637
QY 1382 GCCACTGTGCCCCACAAAGCCCTGGGCGGTACCAATCCTCCCTGCTGCCCTGGGGGC 1441
DB 638 AlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaProGlySer 657
QY 1442 ATGAAGAACCCCGCCAGACCAACCC- -GTCAAGCAGCTCTTCACACAGTGTGTC 1495
DB 658 MetLysSerProThrAspGlnProThrValValLysHisLeuPheThrThrGlyValVal 677
QY 1496 TACGACAGCTTCATCTAAAGCAGCAGTGCATGCGGGAACACACACAGCTGCACCTGAG 1555
DB 678 TyrAspThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGlu 697
QY 1556 CATGCTGCGCGATCCGAGCATCTGGTCCCGCTGCGAGGACAGAGGCTGCTTAGCAAG 1615
DB 698 HisAlaGlyArgIleGlnSerIleIleIleIleIleIleIleIleIleIleIleIle 717
QY 1616 TCCGAGCGGATCCGAGGTCGAAAGCCAGCTAGTAGATGATCCAGACAGTGCACCTGAA 1675
DB 718 CysGluArgIleArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGlu 737
QY 1676 TACCACACCTGCTCTATGAGGACAGTCCCTCAACCGGACAGCTAGACAGCAAGAG 1735
DB 738 TyrHisThrLeuLeuTyrGlyThrSerProLeuAsnArgGlnLysLeuAspSerLysLys 757
QY 1736 TTGCTCGGTCCCATCAGCAGAGATGATGCTGTGCTGCTGCTGGGGGACATCGGGGTG 1795
DB 758 LeuLeuGlyProIleSerGlnLysMetTyrAlaMetLeuProCysGlyGlyIleGlyVal 777
QY 1796 GACAGTGACACCGTGTGGAATGAGATGACCTCCTCCAGTGTGCTGCGGATGCGAGTGGC 1855
DB 778 AspSerAspThrValTrpAsnGluMetHisSerSerSerAlaValArgMetAlaValGly 797
QY 1856 TCCTCTGCTGGAGCTGCGCTTCAAGTGGCTGCGAGGAGCTCAAGAAATGGATTTCGCATC 1915
DB 798 CysLeuValGluLeuAlaPheLysValAlaAlaGlyGluLeuLysAsnGlyPheAlaIle 817
QY 1916 ATCCGCGCCCGCAGGACACACCGCCCGAGGAATCCACAGCATGGGATTCCTCTTCAAC 1975
DB 818 IleArgProProGlyHisHisAlaGluGluSerThrAlaMetGlyPheCysPheAsn 837
QY 1976 TCTGTAGCATCACCCGCAAAACTCTACAGCAGAGATTGAACGTGGGCAAGGTCTCATC 2035

QY 2666 GCAGGTGAGCCGAGAGCCGAAAT-GTGAACGCCATGGCTTCTGTTGGTGGGGCC 2724
 Db 1043 LysCysGluAsnGluGluAlaGluThrValThrAlaMetAlaSerLeuSerValGlyVal 1062
 QY 2725 GAACAGGCCCAAGCTGCGGAGCCGCGGGAACACACAGCCCGCCGAGAGAGCCCATG 2784
 Db 1063 LysProAlaGlu-----LysArgProAspGluProMet 1074
 QY 2785 GAGCAGGAGCTGCCCTG 2802
 Db 1075 GluGluGluProProLeu 1080

RESULT 4
 HD4_HUMAN
 ID HD4_HUMAN STANDARD; PRT; 1084 AA.
 AC P56524; Q9UND6;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Histone deacetylase 4 (HD4) (HA6116).
 GN HDAC4 OR KIAA0288.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99238449; PubMed=10220385;
 RA Grozinger C.M., Hassig C.A., Schreiber S.L.;
 RT "Three proteins define a class of human histone deacetylases related
 to yeast Hda1p.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:4868-4873(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97323006; PubMed=9179496;
 RA Ohara O., Nagase T., Ishikawa K.-I., Nakajima D., Ohira M.,
 Saki N., Nomura N.;
 RT "Construction and characterization of human brain cDNA libraries
 suitable for analysis of cDNA clones encoding relatively large
 proteins.";
 RL DNA Res. 4:53-59(1997).
 RN [3]
 RP REVISIONS TO N-TERMINUS.
 RA Ohara O., Nagase T., Ishikawa K.-I., Nakajima D., Ohira M.,
 Saki N., Nomura N.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=99455038; PubMed=10523670;
 RA Wang A.H., Bertos N.R., Vezmar M., Pelletier N., Crosato M.,
 Heng H.H., Th'ng J., Han J., Yang X.J.;
 RT "HDAC4, a human histone deacetylase related to yeast HDAC1, is a
 transcriptional corepressor";
 RL Mol. Cell. Biol. 19:7816-7827(1999).
 CC -!- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
 THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
 CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
 CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS (BY
 CC SIMILARITY). REPRESSORS TRANSCRIPTION.
 CC -!- SUBUNIT: INTERACTS WITH MEF2C.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -!- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
 CC FAMILY. HD SUBFAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: AFI32607; AAD29046.1; -
 DR EMBL: AB006626; BAA22957.2; ALT_INIT.
 DR Genew; HGNC:14063; HDAC4.
 DR MIM; 605314; -
 DR InterPro; IPR000286; His_deacetylase.
 DR Pfam; PF00850; Hist_deacetyl; 1.
 DR PRINTS; PR01270; HDASUPER.
 KW Hydrolase; Nuclear protein; Transcription regulation; Repressor.
 FT DOMAIN 655 1084
 SQ SEQUENCE 1084 AA; 119069 MW; DF5F30DA9C4295FD CRC64;

Alignment Scores:
 Pred. No.: 1 24e-105 Length: 1084
 Score: 2691.00 Matches: 568
 Percent Similarity: 71.38% Conservative: 118
 Best Local Similarity: 59.11% Mismatches: 195
 Query Match: 50.34% Indels: 81
 DB: 1 Gaps: 23

US-09-502-945-2 (1-2885) x HD4_HUMAN (1-1084)
 QY 2 GAATTCCTCTGTGCAAGTCAAAAGAGAGCCACACAGCGGCGCTCAACCATTCCTCCCA 61
 Db 177 GluPheValLeuAsnLysLysAlaLeuAlaHisArgAsnLeuAsnHisCysIleSer 196
 QY 62 CACACCCCCCAATGCTGG-----GGAGCCACCATTGCTTCTTTGGACGAGATTCCCT 115
 Db 197 SerAspProArgTyrTrpTyrGlyLysThrGlnHisSerSerLeuAspGlnSerPro 216
 QY 116 CCCACAGAGGGCCCCCTGGGAGCGCTCCTCAACAACTGCCTTGGCTGGGCGCTAC 175
 Db 217 ProGlnSer-----GlyValSerThrSerTyrAsnHisProValLeuGlyMetTyr 233
 QY 176 GACAGTCGAGACGACTTCCCTCCGCAAAACAGCGCTCTGAACCCCACTTGAAGTGCCT 235
 Db 234 AspAlaLysAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysLeuArg 253
 QY 236 TCAGGCTAAACAGAGGTGCTGAGCGGAGAGAGTCCCTCTCTGCTGCGCAGGAT 295
 Db 254 SerArgLeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgLysAsp 273
 QY 296 GGGACTGTTATTAGCACACCTTTAAGAGAGAGCTGTTGAGATCACAGTCCCGGCGCTGG 355
 Db 274 GlyProValValThrAlaLeuLysLysArgProLeuAspValThr----- 288
 QY 356 GCGTCGTCGCTGTGTAACAGCGACCCCGCTCCGCGCCCGAGCTCTCCC---AACAGCTCC 412
 Db 289 ---AspSerAlaCysSerSerAlaProGlySerGlyProSerSerProAsnAsnSerSer 307
 QY 413 CACAGCACCATCGCTGAGATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATG 472
 Db 308 GlySerValSerAlaGluAsnGlyIleAlaProAlaValProSerIleProAlaGluThr 327
 QY 473 CTCCTCTCAGCAGCGAGCCCTCCTCTGGACAGTCCCAACAGCTTACAGCTCTACAG 532
 Db 328 SerLeuAlaHisArgLeuValAlaArgGluGlySerAlaAlaProLeuProLeuTyrThr 347
 QY 533 TCTCTCTCTCTCCCAACATCTCCCTAGGGCTGCGAGCGCCACGGTCACTGCACCACTCA 592
 Db 348 SerProSerLeuProAsnIleThrLeuGlyLeuProAlaThr----- 361
 QY 593 CACTTCAGTCCCTCCCGAAGCTGTGCACA-----CAGCAGAGGCGCAGAGGAGCGCC 646
 Db 362 -----GlyProSerAlaGlyThrAlaGlyGlnGlnAspThrGluArgLeuThr 377
 QY 647 CTCAGTCCCTCGGCGAG-----GGTGGCAGCGCTACCGGCAAGTTCATG 691
 Db 378 LeuProAlaLeuGlnGlnArgLeuSerLeuPheProGlyThrHisLeuThrProTyrLeu 397
 QY 692 AGCATCATCTCTATTCTCGCTGCCTGCTGGCGTGGCAGTGGAGGCGAGGAGGAGCC 751
 Db 398 SerThrSer-----ProLeuGluArgAspGlyGlyAla 408

QY 752 CACGGGATGCTCCCTGCTGCACATGCTGCTGTGGAGCAGCGCCGCGCAGCAGC 811
DB 409 ---AlaHisSerProLeuLeuGlnHisMetValLeuLeuGlnProProAlaGlnAla 427
QY 812 ACCCTCAAT-----GCTGTGCCACTCCACGGGAGTCCCACTAGTACGGGT 859
DB 428 ProLeuValThrGlyLeuGlyAlaLeuProLeuHisAlaGlnSer---LeuValGlyAla 446
QY 860 GAACGTGGGCCACACGATCGGGACGCTAGGAGCTCCCGCGGCATCGCCCTGAGC 919
DB 447 AspArgValSerProSer-----IleHisLysLeuArgGlnHisArgProLeuGly 463
QY 920 CGCACTAGTCTCAACCGCTCGCGCAGAGTCCCGAGGCGCTGCACGACTGGTATGCAA 979
DB 464 ArgThrGlnSerAlaProLeuProGlnAsnAlaGlnAlaLeuGlnHisLeuValIleGln 483
QY 980 CAACAGCACCGACGCTCTCTGGAGAAGCAGAG-----CACGACGACTACAG 1027
DB 484 GlnGlnHisGlnGlnPheLeuGlnLysHisLysGlnGlnPheGlnGlnGlnLeuGln 503
QY 1028 CTGGGCAGATCCTCACCAAGACAGGGAGGAGTCCCGCAGCAGCCACACCCACCTGAG 1087
DB 504 MetAsnLysIleProLysProSerGluProAlaArgGlnProGluSerHisProGlu 523
QY 1088 GAGACAGGAGGAGCTGACGAGCAGCAGGAGGAGTCTGCTGGGGAGGAGCCCTG--- 1144
DB 524 GluThrGluGluLeuArgGlnHisGln---AlaLeuLeuAspGluProTyrLeuAsp 542
QY 1145 ACCATGCCCCGGGGGCTCCACAGAGAGTGAGAGCAGACAGAGAGACTGGAGGAGG 1204
DB 543 ArgLeuProGlyGlnLysGluAlaHisAlaGlnAlaGlyValGln---ValLysGlnGlu 561
QY 1205 GACGAGGAAGAGGAGGGAGGAGGAGGAGTTCATCCAGGTAAAGCAGGAGGCGC 1264
DB 562 ProIleGluSerAspGluGluAlaGlu-----ProProArgGluVal 576
QY 1265 GAGAGTGGTCTCAGAGGGGCGCGACTTGGAGGAGCCTGGTGGATACAAAAACTG 1324
DB 577 GluProGlyGlnArgGln---ProSerGluGlnGluLeuLeuPheArgGlnAlaLeu 595
QY 1325 TTCTCAGATGCCACCGCTTTCACACCTTTGACAGGTGTACCAAGCGCCCTCAGCCTGGCC 1384
DB 596 LeuLeuGluGlnGlnArgIleHisGlnLeuArgAsnTyrGlnAlaSerMetGluAlaAla 615
QY 1385 ACTGTGCC-----CACCAAGCCCTGGGCGCTACCCCAATCCTCCCTGCT 1429
DB 616 GlyIleProValSerPheGlyGlyHisArgProLeuSerArgAlaGlnSerSerProAla 635
QY 1430 GCC-----CCTGGGGCATGAAGAACCCCGCAGACCAACCGCTCAAGCAGCTCTTC 1480
DB 636 SerAlaThrPheProValSerValGlnGluProProThrLysPro-----ArgPhe 652
QY 1481 ACCCAAGTGTGTCTACACACGCTTCATGCTAAAGCAGCAGTGCATGTGGGGAACACA 1540
DB 653 ThrThrGlyLeuValTyrAspThrLeuMetLeuLysHisGlnCysThrCysGlySerSer 672
QY 1541 CACGTCAACCTGAGCATGCTGGCCGGATCCAGAGCATGTGGTCCCGGTGCAGGAGACA 1600
DB 673 SerSerHisProGluHisAlaGlyArgIleGlnSerIleTyrSerArgLeuGlnGluThr 692
QY 1601 GGCCTGTTAGCAAGTCCGAGCGGATCCGAGGTGCGCAAGCCACGCTAGATGATCCAG 1660
DB 693 GlyLeuArgGlyLysCysGluCysIleArgGlyArgLysAlaThrLeuGluGluLeuGln 712
QY 1661 ACAGTGCACCTCAATACACACCTCTCTATGGACAGATCCCTCAACCGGCGAGAAG 1720
DB 713 ThrValHisSerGluAlaHisThrLeuLeuTyrGlyThrAsnProLeuAsnArgGlnLys 732
QY 1721 CTAGACAGCAAGATGTGCTCGGTCCCATCATGACGACAGGAAGATGTATGCTGCTCCTGT 1780
DB 733 LeuAspSerLysLysLeuLeuLeuGlySerLeuAla---SerValPheValArgLeuProCys 751

QY 1781 GGGGGCATCGGGTGGACAGTGCACACCGTGTGGAATGAGATGCACCTCTCCAGCTGCTGTG 1840
DB 752 GlyGlyValGlyValAspSerAspThrIleTyrAsnGluValHisSerAlaGlyAlaAla 771
QY 1841 CGATGGCATGGCTGGCTGGAGCTGGCCCTTCAAGGTGGCTGCAGGAGAGCTCAAG 1900
DB 772 ArgLeuAlaValGlyCysValValGluLeuValPheLysValAlaThrGlyGluLeuLys 791
QY 1901 AATGGATTTCCATCATCCCGGCCCGCAGGACACAGCCGAGGAATCCACAGCATGGGA 1960
DB 792 AsnGlyPheAlaValValArgProProGlyHisHisAlaGluGluSerThrProMetGly 811
QY 1961 TTCTGCTTCTCAACTCTGTAGCATCACCGCAAAACTCTACAGCAGAAAGTTGAACGTG 2020
DB 812 PheCysTyrPheAsnSerValAlaValAlaLysLeuLeuGlnGlnArgLeuSerVal 831
QY 2021 GGCAGGTCTCATCTGCTGGAGTGGACATTCACCATGGCAATGGCACCAGCGGCTTC 2080
DB 832 SerLysIleLeuIleValAspTyrAspValHisHisGlyAsnGlyThrGlnGlnAlaPhe 851
QY 2081 TACATGACCCCTCTGTGCTCTACATCTCTGCTATGACATGACAAACGGGAATCTCTTT 2140
DB 852 TyrSerAspProSerValLeuTyrMetSerLeuHisArgTyrAspAspGlyAsnPhePhe 871
QY 2141 CCAGGCTCTGGGGCTCTCTGAAGAGTGGTGGAGGACCGAGGCTGGGGTCAATGTGAAC 2200
DB 872 ProGlySerGlyAlaProAspGluValGlyThrGlyProGlyValGlyPheAsnValAsn 891
QY 2201 GTGCATGGACAGAGGTGGACCCCGCCATTTGGAGACGTGGAGTACCTACAGCCCTC 2260
DB 892 MetAlaPheThrGlyLeuAspProMetGlyAspAlaGluTyrLeuAlaAlaPhe 911
QY 2261 AGGACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2320
DB 912 ArgThrValValMetProIleAlaSerGluPheAlaProAspValValLeuValSerSer 931
QY 2321 GGGTTGATGCTGTGAAGACATCTGCTCTCTGGTGGCTACTCTGTGCTGGTGGTGGTGG 2380
DB 932 GlyPheAspAlaValGluGlyHisProThrProLeuGlyTyrAsnLeuSerAlaArg 951
QY 2381 TGTTCGTCCTGACCGCAGCTGATGACCTGGCAGGGGGCGGGTGGTGGTGGTGGTGG 2440
DB 952 CysPheGlyTyrLeuThrLysGlnLeuMetGlyLeuAlaGlyGlyArgIleValLeuAla 971
QY 2441 CTGAGGAGGAGCCATGACTTACCGCCATCTGTGATGCTCTGAAGCTTGTGTCTCGGCT 2500
DB 972 LeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValSerAla 991
QY 2501 CTGCTCAGTGTAAAGCTGCACGCCCTTGGATGAGGAGTCTTGGCAGCAAAACCCCAAC 2560
DB 992 LeuLeuGlyAsnGluLeuAspProLeuProGluLysValLeuGlnGlnArgProAsnAla 1011
QY 2561 AACGAGTGGCCACGCTAGAAAAGTCTATCGAGATCCAGACAAACACTGGAGCTGTGTG 2620
DB 1012 AsnAlaValArgSerMetGluLysValMetGluIleHisSerLysTyrTyrArgCysLeu 1031
QY 2621 CAGAGTTCGCGCTGGTCTGGCGGCTCTGGCGGCTCTGGAGGGCCCAAGCAGGAGCCCAA 2680
DB 1032 GlnArgThrThrSerThrAlaGlyArgSerLeuIleGluAlaGlnThrCysGluAsnGlu 1051
QY 2681 GAAGCCGAAAT--GTGAACCCATGGCCCTTGTGGTGGGGCCGGAACAGGCCCAAGCT 2739
DB 1052 GluAlaGluThrValThrAlaMetAlaSerLeuSerValGlyValLysProAlaGlu--- 1070
QY 2740 GCGGACCCCGGGGAACACAGCCCCAGCGCCGAGGAGGCCCATGGACGAGGAGCCCTGCC 2799
DB 1071 -----LysArgProAspGluGluProMetGluGluGluPro 1083
QY 2800 CTG 2802
DB 1084 Leu 1084
RESULT 5

HDA9_HUMAN
ID HDA9_HUMAN STANDARD; PRT; 1011 AA.
AC Q9UKV0; Q94845;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Histone deacetylase 9 (HD9) (HD7B) (HD7).
GN HDAC9 OR HDAC7B OR HDAC7 OR KIAA0744.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 4), AND CHARACTERIZATION.
RX MEDLINE=21438017; PubMed=11535832;
RA Zhou X., Marks P.A., Rifkind R.A., Richon V.M.;
RT "Cloning and characterization of a histone deacetylase, HDAC9.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:10572-10577(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [3]
RP SEQUENCE OF 99-650 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=99455038; PubMed=10523670;
RA Wang A.H., Bertos N.R., Vezmar M., Pelletier N., Crosato M.,
RA Heng H.H., Th'ng J., Han J., Yang X.J.;
RT "HDAC4, a human histone deacetylase related to yeast HDAL, is a
RT transcriptional corepressor.";
RL Mol. Cell. Biol. 19:7816-7827(1999).
CC -1- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
CC THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS. ACT
CC AS A TRANSCRIPTIONAL REPRESSOR.
CC -1- SUBUNIT: Interacts with MEF2.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms; 1 (shown here), 2,
CC 3/HD9P and 4/HDAC9a; are produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY. HD SUBFAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC
CC EMBL; AY032737; AAK66821.1; -
CC EMBL; AY032738; AAK66822.1; -
CC EMBL; AB018287; BAA34464.1; -
CC EMBL; AF124924; AAF04254.1; -
CC MIM; 606543; -
CC InterPro; IPR000286; Hist_deacetylase.
CC Pfam; PF00850; Hist_deacetyl; 1.
CC PRINTS; PR01270; HDASUPER.
KW Hydrolase; Nuclear protein; Transcription regulation; Repressor;
KW Alternative splicing.
FT VARSPLIC 487 574
FT VARSPLIC 575 590
FT MISSING (IN ISOFORM 2).
FT PFEPTHTRALSVRQA -> VICKDLAPGVIVKII (IN
FT ISOFORM 3).
FT VARSPLIC 591 1011
FT MISSING (IN ISOFORM 3).
FT GTGLGGYNINIAWTGGLD -> RFISLEPHEFYLSGNCI
FT A (IN ISOFORM 4).
FT

FT VARSPLIC 880 1011 MISSING (IN ISOFORM 4).
FT CONFLICT 99 99 L -> I (IN REF. 3).
FT CONFLICT 437 437 T -> P (IN REF. 3).
FT CONFLICT 644 647 HQCV -> KPNS (IN REF. 3).
SQ SEQUENCE 1011 AA; 111297 MW; 43ED2785E73CD924 CRC64;
Alignment Scores:
Pred. No.: 5,73e-101 Length: 1011
Score: 2579.00 Matches: 540
Percent Similarity: 72.05% Conservative: 107
Best Local Similarity: 60.13% Mismatches: 179
Query Match: 48.24% Indels: 72
DB: 1 Gaps: 19
US-09-502-945-2 (1-2885) x HDA9_HUMAN (1-1011)
QY 2 GAATTCCTCTGTGCGAAGTCA-----AAGGAGGCCACACACGCGCGCTCAACCATTC 55
Db 149 GluPheLeuLeuSerLysSerAlaThrLysAspThrProThrAsnGlyLysAsnHisSer 168
QY 56 CTCCACAGACACCCAAATGCTGG-----GGAGCCACCATGCTCTCTTTGGACAGAGT 109
Db 169 ValSerArgHisProLysLeuTyrThrAlaAlaHisHisThrSerLeuAspGlnSer 188
QY 110 TCCCTCCCGCAGAGCGGCCCTCCCTGAGGCGCTCCCTCTACAACTGCTTTGCGCTGG 169
Db 189 SerProLeuSer-----GlyThrSerProSerTyrLysTyrThrLeuProGly 205
QY 170 CCTACGACAGTCGAGACGACTTCCCTCCGCAAAACAGCCCTCTGAACCACTTGA 229
Db 206 AlaGlnAspAlaLysAspPheProLeuArgLysThrAlaSerGluProAsnLeuLys 225
QY 230 GTCCGTTCAAGGCTAAACAGAAAGTGGTGAGCGGAGAGCAGTCCCTCTCGCTCGC 289
Db 226 ValArgSerArgLeuLysGlnLysValAlaGluArgSerSerProLeuLeuArgArg 245
QY 290 AAGGATGGGACTCTTATTAGCACCTTAAGAGAGAGAGCTGTGAGATCAGACAGTGC 349
Db 246 LysAspGlyAsnValValThrSerPheLysLysArgMetPheGluValThr----- 262
QY 350 CTGGGGCGCTGCTGTTGTAACAGCGACCGCGGCTCGCGGCCCGCCAGCTCTCCCAACAGC 409
Db 263 -----GluSerValSerSerSerSerProGlySerGlyProSerProAsn--- 279
QY 410 TCCACAGACCATCTCTGAGATGCTTACTGGCTCAGTCCCAACATCCCACTAG 469
Db 280 -----AsnGlyProThrGlySerValThrGluAsnGluThrSer 292
QY 470 ATGCTCCCT-----CAGCAGCGAGCGCTCCCTCTG 499
Db 293 ValLeuProProThrProHisAlaGluGlnMetValSerGlnGlnArgLeuLeuHis 312
QY 500 GACAGCTCCCGCAACAGTTACGCTCTACAGCTCTCTCTCTGCCCCAACATCTCCCTA 559
Db 313 GluAspSerMetAsnLeuLeuSerLeuTyrThrSerProSerLeuProAsnIleThrLeu 332
QY 560 GGCTCTCAGCGCCACGGTCTACTGTACCACTCAGCTCCTCCCTCCCGAGCTGTCG 619
Db 333 GlyLeuProAla-----ValProSerGlnLeuAsnAlaSerAsnSerLeuLys 348
QY 620 ACACAGCAGGAGCGCGAGGAGCGGCCCTCCAGTCCCTCGCGCAGGCTGGCAGCTGACC 679
Db 349 GluLysGlnLysCysGlu-----ThrGlnThrLeuArgGlnGlyValProLeuPro 365
QY 680 GGAAGTTC-----ATGAGCACATCTCTATTCTCTGGTGGCTGCTGGGCGCTG 727
Db 366 GlyGlnTyrGlySerIleProAlaSerSerSerHisPro-----HisVal 381
QY 728 GCATCTGAGGGGAGCGGAGCGGCCCGCCAGCGGATGCTCCCTGCTGACGATGCTGTG 787
Db 382 ThrLeuGluGlyLysProAsnSerSerHisGlnAlaLeuLeuGlnHisLeuLeu 401
QY 788 CTGGACAGCGCGCGGAGCAGAGCAGCACCCTCATTTGCT-----GTGCCACTCCACGGG 841

```
Db 402 LysGluGlnMetArgGlnGlnLysLeuValAlaGlyGlyValProLeuHisProGln 421
||||| ||||||| |||:|||| ||||||||||| |||
842 TCCGCACATAGTACGGGTAACTGTGGCCACACAGCATGCGGAGGTAGGCAAGCTCCG 901
||||| ||| |||||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
422 SerProLeuAlaThrLysGluArgIleSerProGlyIleArgGlyThrHisLysLeuPro 441
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
902 CGGCATCGCCCTGAGCGGCACACTCAGTCCCTCAGCGTCCGCGCAGTCCCGCCGCTG 961
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
442 ArgHisArgProLeuAsnArgThrGlnSerAlaProLeuProGlnSer -----ThrLeu 459
962 CAGCAGCTGGTTCATCAACAACAGCAGCAGCAGTTCCTGGAGAACGAAG-----CAG 1015
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
460 AlaGlnLeuValIleGlnGlnHisGlnGlnPheLeuGluLysGlnLysGlnTyroGln 479
1016 CAGCAGCTACAGCTGGGCAAGATCTCACCAGACAGGGGAGCTGCCAGGAGCCCAACC 1075
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
480 GlnGlnIleHisMetAsnLysLeuLeuSerLysSerIleGluGlnLeuLysGlnProGly 499
1076 ACCCACCCTGAGGACAGAGGAGCTGACGGAGCAGGAGGTCTTGTGGGGCAG 1135
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
500 SerHisLeuGluGluAlaGlnGluLeuGln-----GlyAsp 512
1136 GGAGCCCTGACCATGCCCGGAGGCTCCACAGAGAGTGAGCAGCAGAGAACGACCTG 1195
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
513 GlnAlaMetGlnLysAspArgAlaProSerSer---GlyAsnSerThrArgSerAspSer 531
1196 GAGGAGGAGGACGAGAGATGGGAGGAGGAGGAGGAGGATGTCATCCAGGTAAAGAC 1255
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
532 SerAlaCysValAspAspThrLeuGly-----GlnValGlyAlaValLysValLysGlu 549
1256 GAGGAGGCGCAGAGTGTCTCAGGAGGCGCCGCTTGGAGGAGCCGTGCTCGATAC 1315
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
550 GluProValAspSer-----AspGluAspAlaGlnIleGlnGluMetGluSerGlyGlu 567
1316 AAAAACTCTTCAGATGCCCAACCGCTCAACCT-----TTGCAGGTG 1360
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
568 GlnAlaAlaPheMetGlnGlnProPheLeuGluProThrHisThrArgAlaLeuSerVal 587
1361 TACCAGCGCCCTCAGCTGCGCCACTGTG-----CCCCAACCAACCCCTGGCG 1408
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
588 ArgGlnAlaProLeuAlaAlaValGlyMetAspGlyLeuGluLysHisArgLeuValSer 607
1409 CGTACCATCTCCCTCTCCTCCCTCGCGGCGCATGAAGAACCCCCACACCAACCCGTC 1468
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
608 ArgThrHisSerProAlaAlaSerValLeuProHisProAlaMetAspArgProLeu 627
1469 AAGCAGCTCTTCAACCAAGTGTGTCTACGACACGTTTCATGCTAAAGCAGCAGTGCATG 1528
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
628 GlnProGlySerAlaThrGlyIleAlaTyraAspProLeuMetLeuLysHisGlnCysVal 647
1529 TGGCGGAACACACATGACCTGACATGCTGGCGGATCCAGAGCATCTGGTCCCGG 1588
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
648 CysGlyAsnSerThrThrHisProGluHisAlaGlyArgIleGlnSerIleThrPserArg 667
1589 CTCAGGAGACAGCGCTGTAGCAAGTGGCGGATCGGAGTCCGAGGTCCGAAAGCAGCTA 1648
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
668 LeuGlnGluThrGlyLeuLeuAsnLysCysGluArgIleGlnArgLysAlaSerLeu 687
1649 GATGAGATCCAGACATGCTCTGAATACCAACACCTGCTCTATGGGACCATGCCCTC 1708
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
688 GluGluIleGlnLeuValHisSerGluHisHisSerLeuLeuTyroGlyThrAsnProLeu 707
1709 AACCGGCGAGGTACAGCAGCAAGATGCTCGTCCGTCATCAGCCAGAGATGATGCT 1768
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
708 AspGlyGlnLysLeuAspProArgIleLeuLeuGlyAspSerGlnLysPhePheSer 727
1769 GTCTGCTCTTGGGGCATCGGGGTGGAGTGCAGTGCACCGTGTGGAATGAGATGCACTCC 1828
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
728 SerLeuProCysGlyGlyLeuValAspSerAspThrIleThrAsnGluLeuHisSer 747
1829 TCCAGTGTCTGCGCATGGCAGTGGGCTGCTGCTGGAGCTGGCCTTCAAGTGGCTGCA 1888
|||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 748 SerGlyAlaAlaArgMetalValGlyCysValIleGluLeuAlaSerLysValAlaSer 767
1889 GGAGAGCTAAGATGATTGGCTCATCCGCGCCACAGCACACACCGCGGAGGAATCC 1948
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
768 GlyIleLeuLysAsnGlyPheAlaValValArgProGlyHisHisAlaGluGluSer 787
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1949 ACAGCCATGGGATTCTCTTCACTCTGTAGCCATCACCAGCAAACTCTTACAGCAG 2008
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
788 ThrAlaMetGlyPheCysPheAsnSerValAlaIleThrAlaLysTyroLeuArgAsp 807
2009 AAGTTGAACGTGGCAGGCTCTCATCTGAGCTGGAGCATTCACCATGCAATGGCACC 2068
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
808 GlnLeuAsnIleSerLysIleLeuValAspLeuAspValHisHisGlyAsnGlyThr 827
2069 CAGCAGCGCTTCAATGATGACCCCTCTGTCTACATCTCTGTGCTGCTGATGACAAC 2128
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
828 GlnGlnAlaPheTyroAlaAspProSerIleLeuTyroIleSerLeuHisArgTyroGlu 847
2129 GGGAACTCTTTCAGGCTCTGGGCTCTCAAGAGGTGTGGAGGACGAGCGTGGGG 2188
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
848 GlyAsnPhePheProGlySerGlyAlaProAsnGluValGlyThrGlyLeuGlyGluGly 867
2189 TACAATGTGAACGTGGCATGGACAGGAGGTGTGACCCCTTGTGAGGAGCGAGCTGAGTAC 2248
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
868 TyrAsnIleAsnIleAlaThrThrGlyGlyLeuAspProMetGlyAspValGluTyro 887
2249 CTTACAGCCTTCAGCAGGTGTGATGCCATTCACAGGAGTCTCCCTGATGTGGTCT 2308
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
888 LeuGluAlaPheArgThrIleValLysProValAlaLysGluPheAspProAspMetVal 907
2309 CTAGTCTCCGCGGCTTGTGCTGTGTTGAAGACATCTGTCTCTCTGGTGGCTACTCT 2368
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
908 LeuValSerAlaGlyPheAspAlaLeuGluGlyHisThrProLeuGlyGlyTyroLys 927
2369 GTCACCGCCAGATGTTTGGCCACTTGACAGGAGCTGATGACCTGTGAGTGGTGGCGGG 2428
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
928 ValThrAlaLysCysPheGlyHisLeuThrLysGlnLeuMetThrLeuAlaAspGlyArg 947
2429 TGTGCTGGCGCTTGTGCTGTTGAAGTGCAGCCCTTGGATGAGGAGCTTTCAGCAA 2548
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
968 CysValAsnAlaLeuLeuGlyAsnGluLeuGluProLeuAlaGluAspIleLeuHisGln 987
2549 AAGCCCAACATCAGCAGTGGCCACGCTAGAGAAAGTCACTGAGATCCAGCAGC 2602
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
988 SerProAsnMetAsnAlaValIleSerLeuGlnLysIleIleGluIleGlnSer 1005
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
HDA9_MOUSE STANDARD; PRT; 588 AA.
ID HDA9_MOUSE
AC Q99N13; Q9EP72;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone deacetylase 9 (HD9) (HD7B) (Histone deacetylase-related protein) (MEP2-interacting transcription repressor MTR).
GN HDAC9 OR HDAC7B OR HDRP OR MTR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster / NIH;
RA Zhou X., Richon V.M., Rifkind R.A., Marks P.A.;
RT "Cloning of the mouse HDRP cDNA.";
RL Submitted (FEB-2000) to the ENBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH Swiss; TISSUE=Heart;
```

PubMed-11022042;
 Zhang C.L., McKinsey T.A., Lu J.R., Olson E.N.;
 "Association of COOH-terminal-binding protein (CtBP) and
 MEF2-interacting transcription repressor (MITR) contributes to
 transcriptional repression of the MEF2 transcription factor";
 J. Biol. Chem. 276:35-39(2001).
 -|- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
 THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
 HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
 REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS. ACT
 AS A TRANSCRIPTIONAL REPRESSOR.
 -|- SUBUNIT: Interacts with MEF2.
 -|- SUBCELLULAR LOCATION: Nuclear (By similarity).
 -|- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
 FAMILY. HD SUBFAMILY 2.

 THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: AF235053; AAK15027.1; -;
 EMBL: AF324492; AAG48352.1; -;
 MGD; MGI:1931221; Hdac7b.
 KW Hydrolase; Nuclear protein; Transcription regulation; Repressor;
 Alternative splicing.
 FT CONFLICT 120 120 K -> R (IN REF. 2).
 FT CONFLICT 136 136 K -> R (IN REF. 2).
 FT CONFLICT 177 178 MISSING (IN REF. 2).
 FT CONFLICT 523 523 N -> T (IN REF. 2).
 SQ SEQUENCE 588 AA; 65630 MW; D4BDF3774B3D39EA CRC64;

 Alignment Scores:
 Pred. No.: 3 97e-33 Length: 588
 Score: 950.50 Matches: 228
 Percent Similarity: 63.76% Conservative: 57
 Best Local Similarity: 51.01% Mismatches: 113
 Query Match: 17.78% Indels: 49
 DB: 1 Gaps: 17

 US-09-502-945-2 (1-2885) x HDA9_MOUSE (1-588)
 QY 2 GAATTCCTCTGTGCGAGTCA-----AAGAGGCCACACACGCGCGCTCAACCATTC 55
 Db 149 GluPheLeuLeuSerLysSerAlaThrLysAspThrProThrAsnGlyLysAsnHisSer 168
 QY 56 CTCCACAGCAGCCCAATGCTGG-----GGAGCCACCATGCTCTTTGGACACAGAGT 109
 Db 169 ValGlyArgHisProLysLeuTrpTyrThrAlaAlaHisThrSerLeuAspGlnSer 188
 QY 110 TCCTCTCCACAGAGCGCCCTCCCTGGGACGCTCTCCCTACAACTGCTTTGCGCTGG 169
 Db 189 SerProProlSer-----GlyThrSerProSerTyrLysTyrThrLeuProGly 205
 QY 170 CCTACGACAGTGGAGAGCTATCCCTCCCGAAAACAGCCCTCTGACCACTTGAAA 229
 Db 206 AlaGlnAspSerLysAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLys 225
 QY 230 GTGGGTCAGCGCTAAACAGAGGTGGCTGACGGAGAGCAGTCCCTCTCGCTCG 289
 Db 226 ValArgSerArgLeuLysGlnLysValAlaGluArgSerSerProLeuLeuArg 245
 QY 290 AAGGATGGGACTGTTATTACACCTTTAAGAACAGAGAGCTTTCAGATCAGAGTGC 349
 Db 246 LysAspGlyAsnLeuValThrSerPheLysLysArgValPheGluVal----- 261
 QY 350 CCTGGGGCGTGCCTGTGTAAACAGCGACCCGCTCGCGGCCAGCTCTCCCAACAG 409
 Db 262 ---AlaGluSerValSerSerSerSerProGlySerGlyProSerProAsnAn 280


```

Db 638 ArgAlaLeuArgIleLeuIleValAspTrpAspValHisHisGlyAsnGlyThrGlnHis 657
QY 2075 GCGTTCACAAAGACCCCTCTGTCTCTACATCTCTGTGATCGCTATGACACGGGAAC 2134
Db 658 IlePheGluAspAspProSerValLeuTyrValSerLeuHisArgTyrAspArgGlyThr 677
QY 2135 TCTCTTCA-----GGCTCTGGGCTCCTGAAGAGTGTGTGAGGACACAGGCGTGGG 2188
Db 678 PhePheProMetGlyAspGluGlyAlaSerSerGlnValGlyArgAspAlaGlyIleGly 697
QY 2189 TACAATGTGAACGTGGCATGACAGAGGTGTGGACCCGCCATTTGGAGAGCTGGAGTAC 2248
Db 698 PheThrValAsnValProTrpAsnGly-----ProArgMetGlyAspAlaAspTyr 714
QY 2249 CTTACAGCCTTCAGACAGTGGTGTATGCCCATTTGCCACAGAGTCTCACCTGATGCTGGTC 2308
Db 715 LeuAlaAlaTrpHisArgLeuValLeuProIleAlaTyrGluPheAsnProGluLeuVal 734
QY 2309 CTAGTCTCGCGCTGTCTGTCTGAAGACATCTGTCTCTGTGGTGGCTACTCT 2368
Db 735 LeuIleSerAlaGlyPheAspAlaAlaGlnGly-----AspProLeuGlyGlyCysGln 752
QY 2369 GTACCCGCCAGATGTTTGGCCACTGTGACAGGCGAGTGTACCTGGCAGGGGGCGG 2428
Db 753 ValThrProGluGlyTyrAlaHisLeuThrHisLeuLeuMetGlyLeuAlaGlyGlyArg 772
QY 2429 GTGGTGTGGCCCTGAGGAGGCGCATCTGACCGCCATCTGTGATGCTCTGAAGCT 2488
Db 773 IleIleLeuLeuGluGlyGlyTyrAsnLeuAlaSerIleSerGluSerMetAlaAla 792
QY 2489 TGTGTCTGGCTGTCTAGTAAAGCTGAAGCTGACGCCCTTGGATGAGCGAGTCTGACGAA 2548
Db 793 CysThrHisSerLeuLeuGly-----AspProProGlnLeuThrLeuLeuArg 809
QY 2549 AAGCCCAACATCAACGCGTGGCCAGCTAGAACAGTATGAGAACGATCCAGACAAACAC 2608
Db 810 ProProGlnSerGlyAlaLeuValSerIleSerGluValIleGlnValHisArgIlystyr 829
QY 2609 TGGAGCTGTGTGCAGAGTTCGCCCTGTGTGTGGCGCTGCTGCGAGGGGCCCAAGCA 2668
Db 830 Trp-----ArgSerLeuArgLeuSerLysMet 838

RESULT 9
HDAL_SCHPO
ID HDAL_SCHPO STANDARD; PRT; 687 AA.
AC P56523;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Histone deacetylase clr3 (cryptic loci regulator 3).
GN CLR3 OR SPBC800.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA MEDLINE=98429513; PubMed=9755190;
RX Grewal S.I., Bonaduce M.J., Klar A.J.;
RT "Histone deacetylase homologs regulate epigenetic inheritance of
RT transcriptional silencing and chromosome segregation in fission
RT yeast."; 150:563-576(1998).
RL Genetics [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;

```

```

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shipakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN
CC TRANSCRIPTIONAL REGULATION AND CELL CYCLE PROGRESSION.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY. HD SUBFAMILY 2.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF064207; AAD05212.1; -.
DR EMBL; AL391034; CAC01518.1; -.
DR InterPro; IPR000286; Hist_deacetylase.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PR01270; HDASUPER.
KW Hydrolase; Nuclear protein.
SQ SEQUENCE 687 AA; 76792 MW; 6B0E4184A056D899 CRC64;

Alignment Scores:
Pred. No.: 6.34e-23 Length: 687
Score: 705.50 Matches: 154
Percent Similarity: 56.21% Conservative: 86
Best Local Similarity: 36.07% Mismatches: 144
Query Match: 13.20% Indels: 43
DB: 1 Gaps: 10

US-09-502-945-2 (1-2885) x HDAL_SCHPO (1-687)
QY 1385 ACTGTGCCCCACCAAGCCCTGGCGGTACCACTCCCTCCCT-----GCTGCCCTGGG 1438
Db 22 ThrValThrProTrpSerIleLeuLeuThrAsnAsnLysProMetSerGlySerGluAsn 41
QY 1439 GGCATGAGAAGACCCCAAGACCAACCCGCTCAAGCACCTTTCACCAAGTGTGGTCTAC 1498
Db 42 ThrLeuAsnAsnGluSerHisGluMetSerGlnIleLeuLysLysSerGlyLeuCysTyr 61
QY 1499 GACAGCTTCATGCTAAAGCACACAGTGTGATGCGGGGAACACACACAGTGCACCTGAGCAT 1558
Db 62 AspProArgMetArgPheHisAlaThrLeu---SerGluValAspAspHisProGluAsp 80
QY 1559 GCTGCCCGGATCCAGAGCATCTGTGCGGCTGCGAGGAGACAGCGCTCTTAGCAAGTGC 1618
Db 81 ProArgArgValLeuArgValPheGluAlaIleLysLysAlaGlyTyrValSerAsnVal 100

```

```
QY 1619 GAG-----CGATCCGAGGTGCGCAAGCCAGCTAGATGAGATC 1657
Db 101 ProSerProSerAspValPheLeuArgGileProAlaArgGluAlaThrLeuGluGluLeu 120
QY 1658 CAGACAGTGCACCTCT---GAATACCACACCTGCTCTATGGACAGTCCCTCAACCCG 1714
Db 121 LeuGlnValHisSerGlnGluMetTyrAspValThrAsnThrGluLysMetSerHis 140
QY 1715 CAGAAGCTAGACAGCAAGATTGCTCGGTCCATCAGCCAGAGATGATGCTGTGCTG 1774
Db 141 GluAspLeuAlaAsnLeuGluLysIle----- 149
QY 1775 CTTGTGGGGGCATCGGGGTGGAGTGCACCGTGTGGAATGAGATGCACTCTCCAGT 1834
Db 150 -----SerAspSerLeuTyrTyrAsnAsnGluSerAlaPhe 161
QY 1835 GCTGTGCGCATGCGATGGCTCCCTGCTGGAGCTGGCTTCAAGTGGCTGCAGGAG 1894
Db 162 CysAlaArgLeuAlaCysGlySerAlaIleGluThrCysThrAlaValThrGlyGln 181
QY 1895 CTCAAGAATGGATTTCCTATCATCCGCGCCCGCAGGACACCCGCGAGGAATCCAGCC 1954
Db 182 ValLysAsnAlaPheAlaValArgProGlyHisHisAlaGluProHisLysPro 201
QY 1955 ATGGGATCTGCTTCTCAACTCTGTAGCATCACCAGCAAACTCCTCAGCAGGAAGTTG 2014
Db 202 GlyGlyPheCysLeuPheAsnValSerValThrAlaArgSerMetLeuGlnArgPhe 221
QY 2015 -----AAGTGGGCAAGTCTCATCTGCTGGACCTGACATCACCAGCAAGTGCACC 2068
Db 222 ProAspLysIleLysArgValLeuIleValAspTyrAspIleHisHisGlyAsnGlyThr 241
QY 2069 CAGCAGCGCTTCTACATCAGCCCTGCTGCTCTACATCTCTGCTGATCGCTATGACAC 2128
Db 242 GlnMetAlaPheTyrAspAspProAsnValLeuTyrValSerLeuHisArgTyrGluAsn 261
QY 2129 GGGAACTCTTTCCAGGCTCT-----GGGCTCTCTGAAGAGTGTGGTGGAGACCCAGC 2182
Db 262 GlyArgPheTyrProGlyThrAsnTyrGlyCysAlaGluAsnCysGlyGluGlyProGly 281
QY 2183 GNGGGGTACAATGTGACGTGCATGGACGAGGAGTGTGGACCCGCCCATTCGTCGCTG 2242
Db 282 LeuGlyArgThrValAsnIleProTyrSerCysAla-----GlyMetGlyAspGly 298
QY 2243 GAGTACCTTACAGCTTACAGCAGTGGTGTGATGCCATTCGCCAGAGTCTCAGCTGAT 2302
Db 299 AspTyrIleTyrAlaPheGlnArgValValMetProValAlaTyrGluPheAspProAsp 318
QY 2303 GTGGCTCTAGTCTCGCGGGTTGATGCTGTTGAAGGACATCTCTCTCTCTGGGTGGC 2362
Db 319 LeuValIleValSerCysGlyPheAspAlaAlaLagly-----AspHisIleGlyGln 336
QY 2363 TACTCTGTACCGCCAGATGTTTGGCCACTGACAGCAGCTGATGACCTGGCAGGG 2422
Db 337 PheLeuLeuProAlaAlaTyrAlaHisMetThrGlnMetLeuMetGlyLeuAlaAsp 356
QY 2423 GCGCGGTGGTCTGCGCTGGAGGCGCATGATTCACCGCCATCTGTGATGCTCTCT 2482
Db 357 GlyLysValPheIleSerLeuGluGlyGlyTyrAsnLeuAspSerIleSerThrSerAla 376
QY 2483 GAAGCTTGTGCTCGCTCTGCTAGTGTAAAGCTGCAGCCCTCTGGATGAGGAGCTCTTG 2542
Db 377 LeuAlaValAlaGlnSerLeuLeuGlyIleProGlyArgLeuHisThrThrTyrAla 396
QY 2543 CAGCAAAAGCCCAACATCAACGAGTGGCCAGCTAGAGAAAGTCAAGTCAAGATCCAGAGC 2602
Db 397 CysProGln-----AlaValAlaThrIleAsnHisValThrLysIleGlnSer 412
QY 2603 AAACACTGGAGCTGTGTCAG 2623
Db 413 GlnTyrTrpArgCysMetArg 419
```

```
RESULT 10
HDAL_YEAST
ID HDAL_YEAST STANDARD; PRT; 706 AA.
AC P53973;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone deacetylase HDAL.
GN HDAL OR YNL021W OR N2819.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Andre B., Iraqui Houssaini I., Urrestarazu L.A., Vissers S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97121415; PubMed=8962081;
RA Rundlett S.E., Carmen A.A., Kobayashi R., Bavykin S., Turner B.M.,
RA Grunstein M.;
RT complexes that regulate silencing and transcription."
RL Proc. Natl. Acad. Sci. U.S.A. 93:14503-14508(1996).
CC -!- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
CC THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
CC REGULATION AND CELL CYCLE PROGRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY. HD SUBFAMILY 2.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z71297; CAA95883.1;
DR TRANSFAC; T04600;
DR SGD; S0004966; HDAL.
DR InterPro; IPR000286; His_deacetylase.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PR01270; HDASUPER.
DR Hydrolase; Nuclear protein.
SQ SEQUENCE 706 AA; 80069 MW; 4E7069E66D03264D CRC64;
-----
Alignment Scores:
Pred. No.: 1.57e-21 Length: 706
Score: 672.00 Matches: 162
Percent Similarity: 50.40% Conservative: 87
Best Local Similarity: 32.79% Mismatches: 165
Query Match: 12.57% Indels: 80
DB: 1 Gaps: 15
US-09-502-945-2 (1-2885) x HDAL_YEAST (1-706)
QY 1235 GATTGCATCAGGTTAAGGACGAGGAGGAGTGTGCTGAGGAGGGCGCGACTTG 1294
Db 2 AspSerValMetValLysLysGluVal-----Leu 11
QY 1295 GAGGAGCTGTGTGTCATACATAAAACTGTCTCAGATGCCCAACCGCTGCACACTTG 1354
Db 12 GluAsnProAspHisAspLeuLysArgLysLeuGluGluAsnLys----- 26
QY 1355 CAGGTGTACCAAGCGCCCTCAGCCTGGCCACTGTGCCCCACCAAGCCCTGGCGCTACC 1414
Db 27 -----GluGluGluAsnSerLeuSerThr-----SerLysSer 38
QY 1415 CAATCTCTCCCTGCTGCCCTGGGGGATGAGAAGCCCCCGGACCAACCCGCTCAAGCAC 1474
```


QY	1715	CAGAGCTAGACACCAAGAGTTGCTCGGTCCCATCAGCCAGAAGATGATGCTGTGCTG	1774
Db	504	GLuGLuLeuMetGLuGLu-----AlaGLuLysGLuPhe-----	514
QY	1775	CCTTGTGGGGCATCTCGGGGTGGACAGCTGGTGGAGTGGAGTCCCTCTCCAGT	1834
Db	515	-----AsnSerIleTyrLeuThrArgAspThrLeu-----Lys	525
QY	1835	GCTGTGGCGATGGCAGTGGGCTCCCTGCTGGAGAGTGGCC-----TTCAAGAGTGGCT	1885
Db	526	ValAlaArgLysAlaValGLyAlaValLeuLeuGlnSerValAspGLuIlePheGluLysAsp	545
QY	1886	GCAGGAGAGCTCAAGAAATGGATTGCCATCATCCGGCCCCCAGGACACCCAGCGAGAA	1945
Db	546	AlaGLyGln--ArgAsnAlaLeuValIleValArgProGLyHisHisAlaSerAla	564
QY	1946	TCCACAGCCATGGGATTCTCTTTCAACTCTGTAGCATCATCCGCAAAACTCTACAG	2005
Db	565	SerLysSerGLyPheCysIlePheAsnAsnValAlaValAlaLysTyrAlaGln	584
QY	2006	CAGAAGTTGAACGTGGCAAGGTCTCATCTGTGGACTGGGACATTCACCATGGCAATGGC	2065
Db	585	ArgArgHisLysAlaLysArgValLeuIleLeuAspTyrAspValHisHisGLyAsnGLy	604
QY	2066	ACCACAGAGCGGTTCTACAATGACCCCTCTGTGCTCTACATCTCTGTGCATCGCTATGAC	2125
Db	605	ThrGlnGLuIlePheTyrGluAspSerAsnValMetTyrMetSerIleHisArgHisAsp	624
QY	2126	AACGGGAACCTTTTCCAGGCTCTGGGCTCCTCGAA-----GAGGTTGGTGGAGGA	2176
Db	625	LysGLyAsnPheTyrPro---IleGLyGluProLysAspTyrSerAspValGLyGluGLy	643
QY	2177	CCAGCGCTGGGTACAATGTGAACGPGGCATGCAGCAGAGGTGGACCCCCCATCTGGA	2236
Db	644	AlaGLyGluGLyMetSerValAsnValProPheSerGly-----ValGlnMetGLy	660
QY	2237	GACGTGGAGTACCTTACAGCCTTCAGGACAGTGGTGATGCCCATTTGCCACGAGTTCTCA	2296
Db	661	AspAsnGLuTyrGlnMetAlaPheGlnArgValIleMetProIleAlaTyrGlnPheAsn	680
QY	2297	CCTGATGCTGCTACTCTCCCGGGTTTGATGCTGTGTGAAGACATCTGTCTCTCTCTG	2356
Db	681	ProAspLeuValLeuIleThrValLeuGluGLyGlyTyrAsnLeuThrSerIleSerAsn	698
QY	2357	GGTGGCTACTCTCTACCGCCAGATGTTTGGCCACTTGCACGACGAGTGTAGCCCTG	2416
Db	699	GLyGLuTyrLysValThrProGluThrPheAlaLeuMetThrTyrGlnLeuSerSerLeu	718
QY	2417	GCAGGGGCGGGTGTGTGGCCCTGGAGGAGGCCCATGACTTGCACCGCCATCTGTGAT	2476
Db	719	AlaGLyGLyArgIleIleThrValLeuGluGLyGlyTyrAsnLeuThrSerIleSerAsn	738
QY	2477	GCCTCTCAAGCTTGTCTCTCGGCTCTGCTCACT---GTAAAGCTGCAGCCCTTGGATGAG	2533
Db	739	SerAlaGlnAlaValCysGluValLeuGlnAsnArgSerMetLeuArgLeuArgGLu	758
QY	2534	-----CGACTCTTGCAGCAAAAGCCCAACATCAACGACGAGTGGCCACGAGCTA	2578
Db	759	GLuLysGluGlnPheAlaThrLysProGlnLysIleGluSerSerCysIleLysThrIle	778
QY	2579	GAGAAATCATCAGATCCAGACCAACACTGGAGCTGTGTGCAGAGTTTC	2629
Db	779	ArgGluValCysAlaValGlnGlnLysTyrTrpSerIleLeuLysGlyPhe	795
RESULT 12			
Y130_ARCFU			
ID	Y130_ARCFU	STANDARD;	PRT; 359 AA.
AC	O30107;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUL-1998 (Rel. 36, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical protein AF0130.		
GN	AF0130.		

Db 77 -----GlyGlyIle---IleAspPheAspThrAsnIleProValGlyValPheAspArg 93
 QY 1820 ATGCACTCTCCAGTGTGTCGGCATGCGCAGTGGCGTCCCTGCTGGAGCTGGCCTTCAAG 1879
 Db 94 -----AlaLeuLeuAlaAlaGlyGlyAlaIleArgAlaAlaGluAla 107
 QY 1880 GTGGTCGAGGAGAGCTCAAGAAATGATTTGCCATCATCCGGCCCGCCAGGACACAGCC 1939
 Db 108 ValLeuAsnLysGluCysGluAsnAlaPheAlaMetIleArgProGlyHisHisAla 127
 QY 1940 GAGGAATCCAGCAGCATGGATTCCTTCACTCTGAGCCTGAGCAGCAGCCGCAAAA--- 1996
 Db 128 LysProTyrIleGlyAlaGlyPheCysTyrLeuAsnAsnMetAlaIleMetValLysTrp 147
 QY 1997 CTCCTACAGCAGAAATGAGCGTGGCAAGCTCCTCATCCGTCGGAGTGGAGCATCACCAT 2056
 Db 148 LeuLeuLysGlnGlyPheGlu-----ArgIleAlaIleLeuAspTrpAspAlaHisHis 165
 QY 2057 GCATATGACCCAGCAGCGCTTACATGACCCCTCTGCTCTACATCTCTCTGAT 2116
 Db 166 GlyAspGlyThrGlnGluIlePheTyrAsnAspArgValLeuPheIleSerThrHis 185
 QY 2117 CCTATGACACAGGAATCTCTCCAGGCTCTGGGCTCTCTGAAGAGGTGGTGGAGGA 2176
 Db 186 GlnMetPro-----LeuTyrProGlyThrGlyTyrProGluGluCysGlyThrGly 202
 QY 2177 CCAGCGTGGGTACAAATGTGAACGTGGCATGCGACGAGGAGTGTGGACCCCCCATTTGA 2236
 Db 203 LysGlyGluGlyTyrThrValAsnIlePro-----LeuProGlyThrGly 218
 QY 2237 GAGCTGGAGTACTTACAGCCCTTACAGGAGTGGTGGATGCCATTCGCCAGCATTTCTCA 2296
 Db 219 AspGluGlyTyrMetValIleAspGluIleIleGluProValValAsnGluPheLys 238
 QY 2297 CCGTATGTGCTCTAGTCTCCCGCGGTGTGATGCTGTGAAGGACATCTGTCT---CCT 2353
 Db 239 ProGlnPheIleAlaIleSerAlaGlyGlnAsp-----AsnHisPheThrAspPro 255
 QY 2354 CTGGGTGGTACTCTCTACCCCGCAGATGTTTGGCCACTTACAGCAGCATGTGATGCC 2413
 Db 256 IleThrSerLeuAlaLeuProTyrThrAsnLeuGlyIleIleAlaAlaMetAlaGlyPheAspLeu 275
 QY 2414 CTGGCA-----GGGCGCGGTGGTGGTGGCTGGCCCTGGAGGAGGCGCATGCTTG 2461
 Db 276 MetAlaGluLysHisCysAspGlyArgLeuValAlaValLeuGluGlyTyrSerVal 295
 QY 2462 ACC---GCCATCTGATGCCCTCTGAAGCTTGTGCTCGGCTCTGCTAGTCTAAAGCTG 2518
 Db 296 GluGlyAlaLeuProTyrThrAsnLeuGlyIleIleAlaAlaMetAlaGlyPheAspLeu 315
 QY 2519 CAGCCCTTG-----GATGAG 2533
 Db 316 SerAlaIleArgGluProGluAsnTyrLeuProGluLeuLeuTyrPargLysArgAspSer 335
 QY 2534 GCAGTCTTCGACAAAGCCCAACATCAACGAGTGGCCAGTGGCCAGCATAGAGAAGTCACTCGAG 2593
 Db 336 AlaLeuValLysLeuLysHisAsnIle-----GluAspValLysArg 349
 QY 2594 ATCCAGGCAACACTGGAGCTGT 2617
 Db 350 ValHisSerLysTyrTrpLysCys 357
 RESULT 13
 YB94_METH
 ID YB94_METH STANDARD; PRT; 331 AA.
 AC 027262;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MTH1194.
 GN MTH1194.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;

OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN
 RP
 RC SEQUENCE FROM N.A.
 RX STRAIN=Delta H;
 RA MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT delAH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE000887; AAB85683.1; -
 CC InterPro; IPR000286; His_deacetylase.
 CC Pfam; PF00850; Hist_deacetyl; 1.
 CC DR PRINTS; PR01270; HDASUPER.
 CC KW Hypothetical protein; Hydrolase; Complete proteome.
 CC SEQUENCE 331 AA; 36722 MW; 489054F32965EDCF CRC64;
 Alignment Scores:
 Pred. No.: 1.46e-09 Length: 331
 Score: 385.00 Matches: 104
 Percent Similarity: 47.16% Conservative: 54
 Best Local Similarity: 31.04% Mismatches: 131
 Query Match: 7.20% Indels: 46
 DB: Gaps: 10
 US-09-502-945-2 (1-2885) x YB94_METH (1-331)
 QY 1535 AACACACAGTGCACCTCGATGATGCTGGCGGATCCAGACATCTGTGCCGCTGCAG 1594
 Db 13 AsnHisGluGlyHisValGluAsnSerGlyArgThrArgAlaIleLeuArgAlaIleGlu 32
 QY 1595 GAGACAGCCCTGCTAGCAAGTGCAGCGGATCCGAGGTGCGAAAGCCAGCTAGATGAG 1654
 Db 33 SerSerAspLeuSerProArgPhe-----ValGluProGlyMetAlaGlyIleAspAsp 50
 QY 1655 ATCCAGACAGTGCATCTGATATACACACCTGCTCTATGGGACCATGCCCTCAACCGG 1714
 Db 51 IleLeuMetValHisSerSerThrHisVal-----GluTyrLeuGluValPheAla 67
 QY 1715 CAGAAGCTAGACAGCAGCAAGTGTGCTGGTCCCATCAGCCAGAGATGATGTGTGCTG 1774
 Db 61 -----GluTyrLeuGluValPheAla 67
 QY 1775 CCTTGTGGGGCATCGGGGTGGACAGTGCACACCGTGTGGAATGAGATGCATCTCCAGT 1834
 Db 68 GlyArgGlyGlyTyrLeuAspTyrAspThr---TyrMetThrProGluSerPheSer 86
 QY 1835 GCTGTGCCATGGCAGTGGGTGGCTGCTGGAGCTGCGCTTCAAGGTGGTGGTGGAGGAG 1894
 Db 87 ValAlaArgLeuSerAlaGlyGly-----AlaMetLeuAlaAlaGluAla 102
 QY 1895 CTCAGAATGGA-----TTTGCCATCATCCGCGCCCGCCAGCAGCAGCAGGAGATCC 1948
 Db 103 LeuArgAspGlyTyrPheSerTyrSerLeuGlyArgProGlyHisHisAlaThrTyrAsp 122
 QY 1949 ACAGCCATGGGATTCGTCTTCTTCAACTCTGTAGCATCATCCGCAAAACTCTCTACAGCAG 2008

```
Db 123 ArgSerMetGlyPheCysIlePheAsnAsnIleAlaIleAlaIleHisAlaArg 142
QY 2009 AAGTTGAACGTGGCAAGTCTCTGCTGGACTGGGACATTCACCATGGCAATGGCACC 2068
Db 143 AsnLeuGlyValSerArgProLeuValLeuAspPheAspValHisGlyAsnGlyThr 162
QY 2069 CAGCAGGGCTTCAATGACCCCTCTGTCTCTACATCTCTCTGCTGCTATGACAC 2128
Db 163 SerSerIlePheTyArgAspArgAspValMetTyIleSerIleHisGln---AspPro 181
QY 2129 GGGAACTTCTTCCAGGCTCTGGGCTCCTGAAGAGTTGGTGGAGGACCGGCTGGG 2188
Db 182 ArgThrLeuPheProGlyThrGlyPheIleAspGluThrGlySerGlyGluGlyGlu 201
QY 2189 TACAATGTGAACGTGGCATGACAGAGGTGTGGACCCCTGAGAGGTGGAGGTGAC 2248
Db 202 PheAsnLeuAsnIleProMetProArgly-----SerGlyAsnArgGluTy 217
QY 2249 CTTACAGCTTCAGACAGTGGTGATGCCCATTTGCCACGAGTTCTACCTGATGTGTC 2308
Db 218 LeuTrpIleLeuGlyMetIleLeuProAlaValLeuGluGlyPheArgProAspMetile 237
QY 2309 CTAGTCTCCGCGGTTTGAAGTCTTGAAGCATCTGCTCTCTCTGCTGGTGGTACTCT 2368
Db 238 PheValSerAlaGlyPheAspAla-----HisArgArgAspProLeuAlaGluIleMet 255
QY 2369 GTCACCCGCGAGATGTTT-----GGCCACTTGACAGGCGAGTGTACGCCCTGGCA 2419
Db 256 ValAspGluGluPhePheSerTrpIleGlyTrpPheIleHisGln-----Thr 271
QY 2420 GGGGGCGGGTGGTGGCGCTGGAGGAGGCCATGACTTGCAGCGCATCTGTGATGCC 2479
Db 272 GlyLeuProCysThrAlaValLeuGluGlyGlyTyArgProGluAlaLeuGlyArgSer 291
QY 2480 TCTGAAGCTGTGCTCGGCTCTCAGTGTAAAGTGGACGCC 2524
Db 292 AsnIleAlaPheMetArgGlyLeuAspGlyGluGluTyArgGluPro 306

RESULT 14
Y245_SYN3
ID Y245_SYN3 STANDARD; PRT; 304 AA.
AC P27202;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein slr0245.
GN SLR0245.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsumoto A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL: D90900; BAA16709.1; --
DR InterPro: IPR000286; His_deacetylase.
DR Pfam: PF00850; His_deacetyl; 1.
DR PRINTS: PRO1270; HDASUPER.
DR KW Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 304 AA; 33649 MW; 69B37CF7EB6360DC CRC64;

Alignment Scores:
Pred. No.: 1.85e-08 Length: 304
Score: 358.50 Matches: 105
Percent Similarity: 47.80% Conservative: 58
Best Local Similarity: 30.79% Mismatches: 138
Query Match: 6.71% Indels: 41
DB: 1 Gaps: 11

US-09-502-945-2 (1-2885) x Y245_SYN3 (1-304)
QY 1487 AGTGTGCTTACGACACGTTTCATGCTAAAGCACACGATGTCGCGGAACACACAGTG 1546
Db 3 AlaIleIleTySerAlaGluPheLeuArgHisGlu-----ThrGlyProThr----- 18
QY 1547 CACCTGAGCATGCTGCCGATCCAGAGCATCTGTCCCGGTCGACGAGGACAGGCGCTG 1606
Db 19 HisProGluCysProAlaArgLeuThrAlaIleAlaThrAlaLeuArgLys----- 35
QY 1607 CTTAGCAAGTGGCAGCGGATCCGAGGTGCGAAAGCCACGCTAGATGATCCAGACAGTG 1666
Db 36 -----MetProGlyAlaAsnTyLeu 42
QY 1667 CACTCTGAATA-CCACACCCTGCTCTATGGACACGTCCTCAACGGGACAGACTAGA 1725
Db 43 HisTrpGlnLysProSerProValThrTrpAsnLeuAspProTyIleLeuArgCysHis 62
QY 1726 CAGCAAGAAGTGTCTCGTCCCATCAGCCAGAGATGATGCTGCTGCTGTGGGG 1785
Db 63 SerGlnGlu-TyrLeu-----AsnLysLeuAlaLysLeuAlaGluLeuGly 78
QY 1786 CATCGGGGTGGACAGTACACACCTGTGGATGAGATGACACTCTCCAGTGTCTGGCAT 1845
Db 78 yGlySerLeuAspAlaAspThrProValSerPro---GlnSerTyAspValAlaArgLe 97
QY 1846 GGCAGTGGGTGCTGCTGGAGCTGCCCTTCAAGTGGTGCAGGAGACTCAAGATGG 1905
Db 97 uAlaValArgAlaTrpLeuAspGlyValAspHisVal---LeuAsnGlnArgGluAlaVa 116
QY 1906 ATTTGCCATCATCCGGCCCCCAGACACACCGCCAGGAAATCCACAGCATGGGATCTG 1965
Db 116 lPheValLeuAlaArgProGlyHisHisAlaIleArgAsnThrGlyMetGlyPheCy 136
QY 1966 CTTCTTCAACTCTGTAGCATCATCCCAAACTCTTACAGCAGAGTGAACGTGGGCAA 2025
Db 136 sLeuLeuAsnAsnValAlaIleAlaAlaHisTyAlaLeuThrArgProGlyValGluAr 156
QY 2026 GPTCTTCATCTGGTGGACATTCACCATGGCAATGGCACCCAGCAGCGCTTCTACAA 2085
Db 156 gValAlaIleLeuAspTrpAspValHisHisGlyAsnGlyThrGluAlaLeuValAspHi 176
QY 2086 TGACCCCTCTGCTCTACATCTCTCTGCTGCTATGACAAACGGGAACCTCTTTCCAG 2145
Db 176 sAsnProArgIlePheTyCysSerLeuHisGlnPhePro-----CysTyProGlu 193
QY 2146 CTCTGGGCTCTGGAAGAGTGTGGAGGACCGGCTGGGTGATGATGTAAGTGGC 2205
Db 193 yThrGlyAlaAlaGlyAspArgGlyGlnHisAspAsnVal-----LeuAsnIlePr 210
QY 2206 ATGGACAGAGGTGTGGACCCCTTACGAGAGTGGAGTACCTTACAGCTTCAGGAC 2265
Db 210 oLeuLysProGlyGlyAspGlyLysVal-----TyrArgGluAlaPheGluHi 226
QY 2266 AGTGGTGTGATGCCCATTTGCCCGCAGGTTCTCAGCTGATGTTGGTCTCTAGTCTCCGCGGGTT 2325
Db 226 sLysValLeuProPheLeuArgGlnValLysProAspLeuLeuValSerAlaGlyTy 246
```


Tue Mar 25 09:28:36 2003

us-09-502-945-2.rsp

Page 21

Search completed: March 21, 2003, 12:47:14
Job time : 78.3901 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:32:38 ; Search time 51.3908 Seconds
(without alignments)
10793.685 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346

Sequence: 1 ggaattcctcttcgaagt.....aaacaaaagttaaaattt 2885

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cg2_1/USPTO.spool/US0502945/runat_14032003_101059_19127/app_query.fasta_1.10979
-DB=PIR_73 -OPMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0502945 @CGN_1.1.263 @runat_14032003_101059_19127 -NCPU=3
-NO_XLPXY -NO_MMAY -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -FGAPEXT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1831.5	34.3	878	2 T17245	hypothetical prote
2	1034	19.3	796	2 T32425	hypothetical prote
3	781.5	14.6	1063	2 T46284	hypothetical prote
4	771	14.4	1095	2 T13964	probable histone d
5	705.5	13.2	687	2 T43797	probable histone d
6	672	12.6	706	2 S62933	hypothetical prote
7	490.5	9.2	359	1 B62666	acetylpolymine am
8	488	9.1	782	2 T22134	hypothetical prote
9	487.5	9.1	310	1 A70481	acetoin utilizatio
10	438.5	8.2	337	2 AB3440	acetylspermidine d
11	432.5	8.1	359	2 T05998	hypothetical prote
12	420	7.9	311	2 D97450	histone deacetylas
13	420	7.9	311	2 AF2668	deacetylase (impor
14	406.5	7.6	517	2 T27101	hypothetical prote

15	389	7.3	380	2	D83174	probable acetylpol
16	386.5	7.2	498	2	F88359	protein Y51H1A.5 l
17	385	7.2	331	1	C69026	acetylpolymine am
18	374.5	7.0	296	2	G87505	histone deacetylas
19	371.5	6.9	335	1	H71071	hypothetical prote
20	358.5	6.7	304	1	S74557	acetylpolymine am
21	357.5	6.7	327	2	G90139	deacetylase, proba
22	352.5	6.6	338	2	H84173	acetoin utilizatio
23	352	6.6	334	2	B75095	probable histone d
24	348	6.5	343	1	G64366	acetylpolymine am
25	339.5	6.4	306	2	AE1911	hypothetical prote
26	338	6.3	425	2	G86217	protein T27G7.14 l
27	335.5	6.3	389	2	C89958	acetate / acetoin
28	331.5	6.2	387	1	S39643	probable acetylpol
29	319	6.0	320	2	G27762	hypothetical prote
30	312	5.8	461	2	T20163	acetylpolymine am
31	309	5.8	351	2	C90142	hypothetical prote
32	308	5.8	465	2	T23963	hypothetical prote
33	305.5	5.7	389	2	E84054	acetoin dehydrogen
34	303.5	5.7	433	1	S22284	transcription regu
35	303	5.7	480	1	S60381	RPD3 protein homol
36	301	5.6	452	2	S64211	probable transcript
37	297.5	5.6	434	2	T11643	histone deacetylas
38	297	5.6	369	2	F81178	histone deacetylas
39	292.5	5.5	405	2	T40300	hypothetical prote
40	292.5	5.5	839	2	F75518	hypothetical prote
41	291	5.4	385	2	A81926	probable histone d
42	288	5.4	513	2	T01413	BHLF1 protein - hu
43	279.5	5.2	660	1	Q0BE3	probable acetylpol
44	273.5	5.1	344	2	B83605	hypothetical prote
45	268	5.0	507	2	T19067	

ALIGNMENTS

RESULT 1

T17245

hypothetical protein DKF2p586J0917.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T17245

R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, September 1999

A:Reference number: Z18722

A:Accession: T17245

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-878 <KOE>

A:CROSS-references: EMBL:AL117455

A:Experimental source: adult uterus; clone DKF2p586J0917

C:Genetics:

A:Note: DKF2p586J0917.1

Alignment Scores:

Pred. No.:	1-6e-88	Length:	878
Score:	1831.50	Matches:	424
Percent Similarity:	56.03%	Conservative:	124
Best Local Similarity:	43.35%	Mismatches:	249
Query Match:	34.26%	Indels:	182
DB:	2	Gaps:	22

US-09-502-945-2 (1-2885) x T17245 (1-878)

QY 11 TTGTGGAAGTCAAGAGCCACACAGCGGCCTCAACCATTCCTCCACAGACACCCC 70

Db 27 LeuGluArgThrValHisProAsnSerProGlyIleProTyArgThrLeuGluProLeu 46

QY 71 AAATGCTGGGAGCCACCACCATCTCTTTGGACACAGAGTTCCCTCCACAGACGCGCCCC 130

Db 47 GluThrGluGlyAlaThrArgSerMetLeuSerPheLeuProProValProSerLeu 66

QY 131 CCTGGGACGCCCTCCCTCCCTACAACTGCTTTTGCCTGGGCCCTACGACAGTCGACACGAC 190


```
QY 1607 CTTAGCAAGTCGCGAGGATCCGAGGTCGCAAGCCACGCTAGATGATGATCCACAGAGTG 1666
Db 465 AlalaargCysLeuIleLeuProAlaArgProAlaLeuGlySerGluLeuLeuThrCys 484
QY 1667 CACTCT---GAATACCAACACCTCTATGGGACGAGTCCCTCAACCGGACGAAGCTA 1723
Db 485 HisSerAlaGluTyrValGluHisLeuArgThrThrGluLysMetLysThrArgAspLeu 504
QY 1724 GACAGCAAGAAGTTGCTCGGTCCTATCCATGACGACGAGATGATGTCTGCTCCCTTGGG 1783
Db 505 HisArgGlu----- 507
QY 1784 GGCATCGGGTGGAGTGACACGCTGTGGAATGAGATGCACCTCTCAGTGCTGTGGC 1843
Db 508 -----GlyAlaAsnPheAspSerIleTyrIleCysProSerThrPheAlaCysAlaLys 525
QY 1844 ATGCGAGTGGCTGCCCTGCTGAGCTGGCTTCAAGTGGCTGAGGAGAGCTCAAGAAT 1903
Db 526 LeuAlaThrGlyAlaAlaCysArgLeuValGluAlaValLeuSerGlyGluValLeuAsn 545
QY 1904 GGATTTGCCATCCTCCGCGCCCGGACACGCGGAGGATCCACAGCCATGGGATTC 1963
Db 546 GllleAlaValValArgProGlyHisHisAlaGluProAsnAlaAlaCysGlyPhe 565
QY 1964 TGCTTCTTCAACTCTGTAGCCATCACCGCAAACTCTCTACAGCAGAACTTGAACGTGGC 2023
Db 566 CysPhePheAsnSerValAlaValAlaAlaArgHisAlaGln-----IleIleAlaGly 583
QY 2024 -----AAGTCTCTCATCGTGGAGTGGGACATTCACCATGGCAATGGACCCACGAG 2074
Db 584 ArgAlaLeuArgIleLeuIleValAspTrpAspValHisHisGlyAsnGlyThrGlnHis 603
QY 2075 GGTTCCTACATGACCCCTCTGCTGTCTACATCTCTCTGCATCGCTATGACACGGGAAC 2134
Db 604 IlePheGluAspAspProSerValLeuTyrValSerLeuHisArgTyrAspArgGlyThr 623
QY 2135 TTCTTTTCCA-----GGTCTCTGGGCTCTCTGAAGAGTGTGTGGAGGACCGGCTGGG 2188
Db 624 PhePheProMetGlyAspGluGlyAlaSerSerGlnValGlyArgAspAlaGlyIleGly 643
QY 2189 TCATATGGAAGTGGCATGGACAGAGGTGTGGACCCGCCCATTTGGAGAGCTGGAGTAC 2248
Db 644 PheThrValAsnValProTrpAsnGly-----ProArgMetGlyAspAlaAspTyr 660
QY 2249 CTTACAGCCTTCAGCACAGTGTGTGATGCCATTGCCAGAGTCTCTCACTGATGTGTC 2308
Db 661 LeuAlaAlaIleHisArgLeuValLeuProIleAlaTyrGluPheAsnProGluLeuVal 680
QY 2309 CTAGTCTCGCGGGTTTGATGCTGTGTAAGGACATCTGTCTCTCTGGTGGCTACTCT 2368
Db 681 LeuIleSerAlaGlyPheAspAlaAlaGlnGly-----AspProLeuGlyGlyCysGln 698
QY 2369 GTCACGCGCAGATGTTTGGCCATGTCACCGACGCTGATGACCTGGCAGGGGGCGGG 2428
Db 699 ValThrProGluGlyTyrAlaHisLeuThrHisLeuLeuMetGlyLeuAlaGlyGlyArg 718
QY 2429 GTGGTGTGGCTGGAGGAGGCGCATGACTTGACCGCATCTGTGATGCTCTCAAGCT 2488
Db 719 IleIleLeuIleLeuGluGlyGlyTyrAsnLeuAlaSerIleSerGluSerMetAlaAla 738
QY 2489 TGTGTCTCGGCTGTGCTAGTGTAAAGCTGACGCTTGGATGGAGGACGCTTTCGACGAA 2548
Db 739 CysThrHisSerLeuLeuGly-----AspProProGlnLeuThrLeuLeuArg 755
QY 2549 AAGCCCAACATCAACGACGATGCCCGCTAGAGAAGTCAATCGATCCAGACGACAAAC 2608
Db 756 ProProGlnSerGlyAlaLeuValSerIleSerGluValIleGlnValHisArgLysTyr 775
QY 2609 TGGAGCTGTGTGCAGAAAGTTCGCGCTGTCTGGGCGGCTCTGCGGAGGCGCCCAAGCA 2668
Db 776 Trp-----ArgSerLeuArgLeuSerLysMet 784
```

```
QY 2669 GGTGAGACCGAAGAA 2683
Db 785 GluAspLysGluGlu 789
RESULT 5
T43797
probable histone deacetylase (EC 3.5.1.1-) clr3 - fission yeast (Schizosaccharomyces p
N:Alternate names: cryptic loci regulator protein clr3
C:Species: Schizosaccharomyces pombe
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T43797
R:Grewal, S.I.; Bonaduce, M.J.; Klar, A.J.
Genetics 150, 563-576, 1998
A:Title: Histone deacetylase homologs regulate epigenetic inheritance of transcriptio
A:Reference number: 225685; MUID:98429513; PMID:9755190
A:Accession: T43797
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-687 <GRE>
A:Cross-references: EMBL:AF064207; NID:94159998; PIDN:AAD05212.1; PID:94159999
A:Experimental source: strain 972
C:Genetics:
A:Gene: clr3
A:Map position: 2
C:Keywords: hydrolase; transcription regulation
Alignment Scores:
Pred. No.: 1,36e-29 Length: 687
Score: 705.50 Matches: 154
Percent Similarity: 56.21% Conservative: 86
Best Local Similarity: 36.07% Mismatches: 144
Query Match: 13.20% Indels: 43
DB: 2 Gaps: 10
US-09-502-945-2 (1-2885) x T43797 (1-687)
QY 1385 ACTGTGCCCCCACCAGCCCTGGGCGCGTACCACATCTCCCTCT-----GCTGCCCTGGG 1438
Db 22 ThrValThrProTrpSerIleLeuLeuThrAsnAsnLysProMetSerGlySerGluAsn 41
QY 1439 GGCATGAAGAACCCCCAGACACCCGTCACACACCTCTTCCACCAAGTGTGGTCTAC 1498
Db 42 ThrLeuAsnAsnGluSerHisGluMetSerGlnIleLeuLysLysSerGlyLeuCysTyr 61
QY 1499 GACAGCTTCATGCTAAAGCACACACAGTGTGCGGGAACACACACGTCACCCCTGAGCAT 1558
Db 62 AspProArgMetArgPheHisAlaThrLeu---SerGluValAspAspHisProGluAsp 80
QY 1559 GCTGGCGGATCCAGAGCATCTGTCCCGCTGCGGCTGAGAGACAGCCCTCTTAGCAAGTGC 1618
Db 81 ProArgArgValLeuArgValPheGluAlaIleLysLysAlaGlyTyrValSerAsnVal 100
QY 1619 GAG-----CGGATCGAGGTGCGAAAGCCACGCTAGATGAGATC 1657
Db 101 ProSerProSerAspValPheLeuArgIleProAlaArgGluAlaThrLeuGluGluLeu 120
QY 1658 CAGACAGTGCACCTCT---GAATACACACACCTGCTCTATGGGACCACTGCCCTCAACCGG 1714
Db 121 LeuGlnValHisSerGlnGluMetTyrAspArgValThrAsnThrGluLysMetSerHis 140
QY 1715 CAGAAGCTAGACAGAAGTTGCTCGGTCCCATCAGCCAGCAAGATGTATGCTGTGCTG 1774
Db 141 GluAspLeuAlaAsnLeuGluLysIle----- 149
QY 1775 CCTGTGGGGGCATCGGGGTGGACAGCTGACACCTGTGGAAATGAGTGCACCTCTCCAGT 1834
Db 150 -----SerAspSerLeuTyrTyrAsnAsnGluSerAlaPhe 161
QY 1835 GCTGTGCGCATGGCAGTGGCTGCTGCTGGAGCTGCCCTTCAAGGTGGCTGCGAGGAG 1894
Db 162 CysAlaArgLeuAlaCysGlySerAlaIleGluThrCysThrAlaValThrGlyGln 181
QY 1895 CTCAGAATGATTTGCCATCATCCGCGCCCCCAGGACACACCGCCGAGGAATCCACAGCC 1954
```

Db	182	VallyAsnAlaPheAlaValValArgProProGlyHisHisAlaGluProHisLysPro	201
QY	1955	ATGGATTCCTGCTTCTCAACTCTGTAGCCATCATCCGCAAACTCCTACAGCAAGTTG	2014
Db	202	GlyGlyPheCysLeuPheAsnAsnValSerValThrAlaArgSerMetLeuGlnArgPhe	221
QY	2015	-----AACCTGGCAGGTCCTCATCTGGAGCTGGACATTCACCATGGCAATGGCACC	2068
Db	222	ProAspLysIleLysArgValLeuIleValAspTrpAspIleHisHisGlyAsnGlyThr	241
QY	2069	CAGCAGCGCTTCAACATCACCCCTCTGTGCTCTACATCTCTCTCATCGCTATGACAAC	2128
Db	242	GlnMetAlaPheThrAspAspProAsnValLeuTyrValSerLeuHisArgTyrGluAsn	261
QY	2129	GGGAACTCTTCCAGGCTCT-----GGGGCTCTGAGAGGTTGGTGGAGGACCAAGC	2182
Db	262	GlyArgPheTyrProGlyThrAsnTyrGlyCysAlaGluAsnCysGlyGluGlyProGly	281
QY	2183	GTGGGTGACAATCTGAACGTGGCATGGACGGAGGTGTGGACCCCCCATTTGGAGACGTG	2242
Db	282	LeuGlyArgThrValAsnIleProItrSerCysAla-----GlyMetGlyAspGly	298
QY	2243	GAGTACCTTACAGCCTTCAGGACAGTGGTGTAGCCCATTTGCCACAGCTTCTCACCTGAT	2302
Db	299	AspTyrIleTyrAlaPheGlnArgValValMetProValAlaTyrGluPheAspProAsp	318
QY	2303	GTGGTCTAGTCTCCCGGGTTGTATGCTGTGTGAAGGACATCTGTCCTCTGGGTGGC	2362
Db	319	LeuValIleValSerCysGlyPheAspAlaAlaAlaGly-----AspHisIleGlyGln	336
QY	2363	TACTCTCTACCCGCCAGATGTTTGGCCACTTGCACCGCAGCTGTAGTACCTGGCAGGG	2422
Db	337	PheLeuLeuThrProAlaIaTyrAlaHisMetThrGlnMetLeuMetGlyLeuAlaAsp	356
QY	2423	GGCCGGTGGTGTGTCGCCCTGGAGGAGGCCATGACTTGCAGCCGCATCTGTGTATGCCTCT	2482
Db	357	GlyLysValPheIleSerLeuGluGlyTyrAsnLeuAspSerIleSerThrSerAla	376
QY	2483	GAAGCTTGTCCTCGGCTCTGCTCAGTGTAAAGCTGCAGCCCTTGGATGAGGCAGTCTTG	2542
Db	377	LeuAlaValAlaGlnSerLeuLeuGlyIleProProGlyArgLeuHisThrThrTyrAla	396
QY	2543	CAGCAAAAGCCCAACATCAACGACGTGGCCGCTAGAGAAGTCATCGAGATCCAGAGC	2602
Db	397	CysProGln-----AlaValAlaThrIleAsnHisValThrLysIleGlnSer	412
QY	2603	AAACACTGGAGCTGTGTGCAG	2623
Db	413	GlnTyrTrpArgCysMetArg	419

RESULT 6
S62933 hypothetical protein YNL021W - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N2819
C:Species: Saccharomyces cerevisiae
C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 16-Dec-1998
C:accession: S62933
R:Andre, B.; Iraqui Houssaini, I.; Urrestarazu, L.A.; Vlissers, S.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62920
A:Accession: S62933
A:Molecule type: DNA
A:Residues: 1-706 <AND>
A:Cross-references: EMBL:271297; NID:g1301850; PID:e239664; MIPS:YNL021W
A:Experimental source: strain S288C

Alignment Scores:					
Pred. No. :	7-73e-28	Length:	706		
Score:	672.00	Matches:	162		
Percent Similarity:	50.40%	Conservative:	87		
Best Local Similarity:	32.79%	Mismatches:	165		
Query Match:	12.57%	Indels:	80		
DB:	2	Gaps:	15		
US-09-502-945-2 (1-2885) x S62933 (1-706)					
QY	1235	GATTGCATCCAGTTTAAAGACGAGGGCGCAGACTGTGCTGAGGAGGGGCCCACTTG	1294		
Db	2	AspSerValMetValIlysylGluVal-	-Leu	11	
QY	1295	GAGGAGCGTGGTCTGGATACAAAACACTGTTCAGATGCCAACCGCTCCAACCTTTG	1354		
Db	12	GluAsnProAspHisAspleuIysargIyLysleuGluGluAsnLys-----	26		
QY	1355	CAGGTGTACAACGGCCCCCTCACGCTGGCCACTGTGCCACCACCAAGCCCTGGCCGTATCC	1414		
Db	27	-----GluGluGluAsnSerLeuSerThrThr-----	SerLysSer	38	
QY	1415	CAATCCTCCCCTGCTCCCTCGGGGCATGAAGAACCCTCAGACCAAACCGCTCAAGCAC	1474		
Db	39	LysArgGlnValIleValProValCystMetProLysIleHISrYSrPr-oLeuLys---	57		
QY	1475	CTCTTCACCAACAAGTGTGCTACGACACGTTTCATGCTAAGACCAACCATGTCATGTGCGG	1534		
Db	58	-----ThrGlyLeuCysTyrrAspValArgMetArgTyrHisAlaIysIlePheThr	74		
QY	1535	AAC-----ACACACGTGCACCCTCAGAGATGCTGGCCGCGATCCAGACATCTCG	1582		
Db	75	SerTyrPheglutryIleAspProHisProGluAspProArgIleTyrArgIleTyr	94		
QY	1583	TCCGCGGTGCAGGACAGCGCTCTTAGCAAG-----	1615		
Db	95	LysileLeuAlaGlusnGlyLeuileasnspprothrLeuSerglylvalaspserLeu	114		
QY	1616	-----TGCGAGGGATCCGAGTGCAGAACCGCTAGATGAGATCCAGACAGTG	1666		
Db	115	GlyAspLeuMetLeuIysileproValArgAlaIalathrsGluGluIleLeuGluVal	134		
QY	1667	CATCTGAATAACCAC---ACCCTGCTCTATGGGACAGTCCCTCAACCGCGCAAGAAGTA	1723		
Db	135	HISThrLysGluHisLeuGluPheIleGluSerThrGluLysMetSerArgGluGluLeu	154		
QY	1724	----GACAGCAAGAAGTGTGCTCGGTCCCATCAGCAGAGATGTATGCTGTGTCGCCT	1777		
Db	155	LeuLySGlutrhGluLys-----	160		
QY	1778	TGTGGGGGATCGGGGTGGACAGTGCACCGTGTGGAAATGAGATGCCTCCCTCAGTGCCT	1837		
Db	161	-----GlyaspserValtyrPheasnspserTyrAlaSer	173		
QY	1838	GTCCGATGGCAGTGGGTGCCTGTGGAGTGGCCCTCAAGTGGCTCGAGGAGAGCTC	1897		
Db	174	AlaArgLeuproCysglycylAlalailegLuacysIysalaValGluGlyArgval	193		
QY	1898	AAGAATGGATTGGCATCATCGGCCCGCCAGGACACACCGCCGAGNATCCACGCCATG	1957		
Db	194	LysAsnserLeuAlaValaValArgproGlyHisIhisAlaGluProGlnAlaLaGly	213		
QY	1958	GGATTCTGCTCTTCAACTCTGTAGCATCACCCAAAA-----CTCCTACAGCAGAAG	2011		
Db	214	GlyPheCysLeuPheserAsnValalaValalaIylsasnlleuIylsasntyrPro	233		
QY	2012	TTGAACGTGGGAAGGTCTCTCATCTGGACTGGGACATTACCATGGCAATGGCACCCAG	2071		
Db	234	GluSerValArgIleWetilleuAsprTrpAspileHisIGlysnGlythrGln	253		
QY	2072	CAGGCGTTTACAAANTGACCCCTCTGTGCTCTACATCTCTTCGATCGCTATGACACGGG	2131		
Db	254	LysSerPheTyrGlnAspGlnValLeutyrrValSerLeuthisArgpheGluMetGly	273		

Db 336 AlaLeuValLysLeuLysHisAsnIle-----||| ||| |||
 QY 2594 ATCCAGACCAACACTGGAGCTGT 2617
 Db 350 ValHisSerLysTyrTrpLysCys 357
 RESULT 8

T22134
 hypothetical protein F436.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T22134

R:Swingburne, J.

submitted to the EMBL Data Library, July 1995

A:Reference number: Z19522

A:Accession: T22134

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-782 <WIL>

A:Cross-references: EMBL:Z50070; NID:e1519046; PIDN:CAA90401.1; GSPDB:GN00020; CESP:F436

A:Experimental source: clone F4366

C:Genetics:

A:Gene: CESP:F4366.4

A:Map position: 2

A:Introns: 81/1; 121/2; 160/1; 204/3; 236/3; 274/3; 324/3; 489/1; 551/2

Alignment Scores:

Pred. No.: 3,32e-18 Length: 782
 Score: 486.00 Matches: 172
 Percent Similarity: 40.78% Conservative: 100
 Best Local Similarity: 25.79% Mismatches: 263
 Query Match: 9.13% Indels: 132
 DB: 2 Gaps: 19

US-09-502-945-2 (1-2885) x T22134 (1-782)

QY 899 CCGGGGATCGGCCCTGAGCCGACCTAGTCTCACCCTGCGCAGAGTCCCGAGGCC 958
 Db 17 ProThrSerSerProSerAlaValPheProProLeuHisIlePheProHis 36
 QY 959 CTCGAG-----CAGCTGTCTATGCAACAAACAGCAGCAGCAG 994
 Db 37 MetGlnGlnTyrPheLeuAsnIleMetGlnLeuValGlnGlnGlnTyrGlnGln 56
 QY 995 TTCCTGGAG-----AAGCAGAAGCAGCAGCAG 1021
 Db 57 ArgLeuSerSerGlySerProValAsnSerAspGlnAlaSerSerArgSerGluSer 76
 QY 1022 CTCAGCTGGGCAAG---ATCCTCACCAAGACAGAGGGAGCTGCCAGGCACCCACC 1078
 Db 77 ProGlnThrLysLysLeuLeuLeuAspGluAspGlyCysAlaThrCysSerValCysThr 96
 QY 1079 CAC-----CCTGAG---GAGACAGAGGAGGAGCTGACGAGCAGCAGGAGTCTTGCTG 1129
 Db 97 GluLysValProGluAlaGlnTyrAsnSerHisIleGluLeuGluLysGluArgLeuIle 116
 QY 1130 GGGGAGGAGCCCTGACCATGCCCGGAGGGCTCCACAGAGTGTAGACACACAGGAA 1189
 Db 117 SerTyrIleThrLeuThrLysGluLysArg---GluArgAspArgGluAsnThrProAsp 135
 QY 1190 GACCTGGAGGAGGAGCAGGAGGAGAG---1216
 Db 136 HisPheAspGlnLysArgLysArgGluLeuLeuLeuGlnArgIleArgAsnAsnGlnAsn 155
 QY 1217 -----GATGGGAGGAGGAGGAGGATTCATCCAG-----GTT 1249
 Db 156 LysArgGlnSerLeuLysArgGlyProGlnLeuValArgAspCysLeuThrProPheSer 175
 QY 1250 AAGGACGAGGAGGAGGAGTGTGTGAGGAGGGGCCACCTTGGAGGAGCCTGTGCT 1309
 Db 176 ArgGlnSerAsnAspGluThrGlySerSerGluSerProAspMetLysLysAspGlu 195

QY 1310 GGATACAAAACACTGTTCTCAGATGCCACACCG-----1342
 Db 196 PheTyrMetLysCysThrThrCysHisGlnProCysSerTyrAlaIleValMetSerAla 215
 QY 1343 -----CTGCAACCT 1351
 Db 216 PheAspArgProLysCysGlnIleCysPheAspLeuValArgAlaAlaValLeuAsnSer 235
 QY 1352 TTGCAGGTGTACCAAGGCCCTCAGCCTGGCCACT-----GTGCC 1393
 Db 236 AlaSerLeuTyrLysAsnSerLeuLeuLeuPheSerGluLysAsnTyrIleGluIleAsn 255
 QY 1394 CACCAAGCCCTGGCGTACCCCAATCTCCCTGCTGCCCTGGGCGCATGAGAGACCCC 1453
 Db 256 AsnSerLeuLeuLysSerIleLysHisSerHisArgLeuMetAsnIleGlnLysAsnMet 275
 QY 1454 CCAGACCAACCCGCTCAAGCACCTCTTCCACCAAGTGTGCTCAGCAGCTTCATGCTA 1513
 Db 276 SerSerLysProPheGlyTyr-----ValTyrAspGluArgMetLeu 289
 QY 1514 AACACACAGTGCATGCGGGGAACACACAGCTGCACCCCTGAGCATGCTGCCGGATCCAG 1573
 Db 290 GlyHisGluCysLysTyrAspSerThrMetAla-----GluCysProLysArgMetLys 307
 QY 1574 AGCATCTGGTCCCGCTGCAGGAGACAGCGCTGTAGCAAGTGCAGCGGATCCAGGT 1633
 Db 308 LeuIleTyrGluArgLeuGlnLysAspLysLeuLeuAspGlyAlaValLysIleAspAla 327
 QY 1634 CGCAAGCCACGCTAGATGAGATCCAGACAGTGCATCTGAATACCACACCCCTGCTCTAT 1693
 Db 328 ArgGluAlaAsnAspSerGluIleArgLeuAsnHisProGln-----341
 QY 1694 GGGACCATGCTCCCTCAACCGGCGAGAGCTAGACAGCAAGTGTGCTCGGTCCTCAGC 1753
 Db 342 -----GlnLeuIleAspGluIleValSerLeuAsnSerThrGlu 354
 QY 1754 CAGAAGATGTATGCTGTGCTGCTGCTGCGGGCATCGGGTGGACAGTGCACCGGTGCG 1813
 Db 355 Lys-----CysGluGluTyrCysLysAspLysGluIleLeuTrp 367
 QY 1814 AATGAGATGACATCTCCAGTGTGTCGATGGCAGTGGCTGGCTGCTGGAGTGCC 1873
 Db 368 ThrCysGluAsnThrAsnGluAlaAlaArgValAlaIleGlySerIleGluLeuVal 387
 QY 1874 TTCAAGGTGCTGCTCAGGAGCTCAAGAATGATTTGCCATCATCCGGCCCCCAGGACAC 1933
 Db 388 LysAlaAlaMetGluGlyLysIleHisAsnGlyPheAlaMetValArgProGlyHis 407
 QY 1934 CAGCGCGAGGAATCCACAGCCATGGGATCTGCTTCTTCAACTCTGTAGCCATCCGCA 1993
 Db 408 HisSerTyrGlyLysValAlaGlnGlyTyrCysIlePheAsnAsnValAlaIleAla 427
 QY 1994 AAATCTCTACAGCAAGTGAACGTGGCGAGGTCTCTCATCTGTGGACTGGACATTCAC 2053
 Db 428 LysTyrAlaIleGluLysLeuGluValLysValAlaIleValAspPheAspTyrHis 447
 QY 2054 CATGGCAATGGCACCAGCAGCGGTCTCAATGATGACCCCTCTGTCTCTACATCTCTG 2113
 Db 448 AlaGlyAsnGlyThrTyrGlnSerValLysAspGluSerArgIleHisLeuThrSerPhe 467
 QY 2114 CATCGTATGACAAACGGAACTTCTTTCAGGCTCTGGGGCTCTGAGAGGTGGTGA 2173
 Db 468 HisGlyHisPheGlyAlaPheTrpProPheSer-----479
 QY 2174 GGACCGAGCGTGGGTACAATGTGAAC-----GTGCATGGACAGA 2215
 Db 480 ---ArgAspTyrAspTyrAlaThrAsnSerGlnAsnThrLeuPheValProLeuAsnGly 498
 QY 2216 GGTGTGGACCCCTTTCAGGACGTGGAGTACCTTACAGCCTTCAGACAGCTGTGTATG 2275
 Db 499 ThrMetAsn-----SerGluGlyAspTyrValSerValPheHisValLeuLeu 515
 QY 2276 CCATTGGCCACGAGTTCTCACTGATGTGGTCTTAGTCTCCCGGGGTTTTCATGCTGTT 2335

```

||||| 516 PrometLeuGluGlnPheGlnProGluLeuLeuLeuSerAlaGlyPheAspSerGly 535
QY 2336 GAAGGACATCTGCTCCCTGCTGGGTGCTGCTACCGCCAGATGTTTGGCCACTTG 2395
Db 536 TyrTyraSpValMetMetGluPheGlyGlnGlyValLysAlaHisGlyTyrGlyHisMet 555
QY 2396 ACCAGGAGCTGATCACCCTGCAGGGGCGGTGGTGGCTGGCCCTGGAGGAGGCAT 2455
Db 556 AlaArgLeuLeuSerGlnIleCysProGlyLysIleLeuAlaIleLeuGluGlyGlyTyr 575
QY 2456 GACTTGACCGCCATCTGTGATGCTCTCAAGCTGTGCTGGCTCTGCTCAGTGTAAAG 2515
Db 576 HisProTyrAsnTyrThrGluSerAlaSerMetMetValArgGlyLeuLeuAsnLeuPro 595
QY 2516 CTGACGCCCTGGATGAGGAGCTGTTGCGCAAGAACCCACATCAACGCGAGTGGCCAG 2575
Db 596 LeuProArgLeuAsp-----lleProGluArgIleSerGlyAlaLeu 609
QY 2576 CTAGAGAAAGTCATCGAGATCCAGAGCAACACACTGGAGCTGTGTGCAGAAAGTTCGCCGCT 2635
Db 610 LeuGluThrThrTrpAsnIleLeuAsnHisHisSerGluTrpTyrProLys----- 626
QY 2636 GCTCGGCGCGTCCCTCGCA 2656
Db 627 ---LeuGlyGluArgLeuLys 632

RESULT 9
A70481
acetoin utilization protein - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: A70481
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'
V. Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID: 98196666; PMID: 9537320
A:Accession: A70481
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-310 <AQP>
A:Cross-references: GB:AE000773; GB:AE000657; NID:g2984313; PIDN:AAC07842.1; PID:g298432
A:Experimental source: strain VF5
C:Genetics:
A:Gene: acu2
A:Superfamily: acetylpyrimidine aminohydrolase; RPD3/acuC homology
F:6-303/Domain: RPD3/acuC homology <RAH1>

Alignment Scores:
Pred. No.: 3 35e-18 Length: 310
Score: 487.50 Matches: 116
Percent Similarity: 50.87% Conservative: 60
Best Local Similarity: 33.53% Mismatches: 125
Query Match: 9.12% Indels: 45
DB: 1 Gaps: 8

US-09-502-945-2 (1-2885) x A70481 (1-310)
QY 1493 GTCTACGACAGTTCATGCTAAAGCACCAGTGCATGTGCGGGAACACACACATGACCCCT 1552
Db 7 IleTyrAspAspIleTyrLeuLysHis-----AspTrpProGluHisPro 21
QY 1553 GAGCATGCTGGCGGATCCAGGACATCTGTCCTCCGCTGCAGGAGACAGCCCTGCTAGC 1612
Db 22 GluAsnLysAsnArgLeuLeuSerIleLeuGluHisValGluLysSerGlyIleLysLys 41
QY 1613 AAGTGGCGGATCCGAGGTCCCAAGCCACCTAGATGAGATCCAGACAGTGCACCTCT 1672
Db 42 AlaLeuIleAspValLysProArgAlaLysValGluGluValAlaLeuAsnHisAsp 61
QY 1673 GAATACACACACCTGCTCTATGGACCACTGCCCTCAACCGGCAAGAGTAGACAGCAAG 1732
Db 1732

```

```

Db 62 ProAlaTyrIle----- 65
QY 1733 AAGTGTGCTCGTCCCATCACCAGAAAGATGTATGCTGTGCTGCTTGTGGGGCATCGGG 1792
Db 66 -----GlnGluIleHisAspPheCysLysSerGlyGlyTyr 78
QY 1793 GTGACAGTGCACACCTGTGGAATGAGATGACACTCCTCCAGTCTGTGGCAGTGGCAGTG 1852
Db 79 LeuAspProAspThr---TyrAlaThrProAspThrTyrAspValAlaLeuTyrAlaVal 97
QY 1853 GGCTGCTGCTGAGTGGCTTCAAGTGGCTGCAGGAGCTGCAGAGAGCTCAAGATGATTTGCC 1912
Db 98 GlyGlyValLeuGluGlyIleAspArgIleLeuSerGlyGluLeuAspArgAlaPheCys 117
QY 1913 ATCATCGCGCGCCAGGACACCGAGGAATGCACACGCAATGGATTCCTGCTTCTTC 1972
Db 118 AlaValArgProGlyHisHisAlaGluTyrAlaLysAlaMetGlyPheCysIlePhe 137
QY 1973 AACTCTAGTCATCACCAGCAAACTCTACAGCAGAAAGTTGAACGTGGGCAAGTCTC 2032
Db 138 AsnAsnValAlaIleGlyAlaHisTyrLeuArgLysIleLysGlyValAsnLysValPhe 157
QY 2033 ATCGTGACTGGGACATTCACCATGGCAATGGCACCAGGCGGTCTACATGACCCC 2092
Db 158 IleIleAspPheAspAlaHisHisGlyAsnGlyThrGlnLysSerPheTyrGluAspAsp 177
QY 2093 TCTGTGCTCTACATCTCTCTGCTGCTATGACAACGGGAATCTTCTCCAGGCTCTGGG 2152
Db 178 ThrValPheTyrPheSerThrHisGluTyrPro-----PheTyrProGlyThrGly 194
QY 2153 GCTCCTGAAGAGGTGTGGGAGGACCGGCTGAGTGTGCAATGTGAACGTGGCATGACA 2212
Db 195 SerGluAspGluArgGlyAlaGlyLysGlyTyrGlyTyrThrTyrAsnValProMetSer 214
QY 2213 GGAGGTGTGACCCCTCCATGGAGCTGAGTGTACCTTACAGCCTTCAGGACAGTGGT 2272
Db 215 AlaGlyAla-----GlyAspAspGluTyrIleProValTyrGluLysLeuLeu 230
QY 2273 ATGCCCATGCCCCAGAGTCTCTACCTGCTGCTGCTAGTCTCCGCGGGTGTGATGCT 2332
Db 231 ProGluLeuMetAsnArgPheSerProAspPheValLeuValSerAlaGlyTyrAspLeu 250
QY 2333 GTTGAAGGACATCTGCTCTCTGCTGGGTGCTACTCTGTACCCGCCAGATGTTTGGCCAC 2392
Db 251 HisArgAsp-----AspProLeuThrTyrLeuAsnValSerAsnGluGlyValArgGln 268
QY 2393 TTGACCGAGGACGTGATGACCTGGCA-----GGGGCGCGGTGGTGGTGGCCCTG 2443
Db 269 IleValArgAsnIleIleLysThrSerAspGluLeuAsnAlaProValLeuPheAlaLeu 288
QY 2444 GAGGAGGCCCATGACTTGACCGCCATCTGTGATGCTCTGAAGCTTGTCTCGGCTCTG 2503
Db 289 GluGlyGlyTyrAsnLeuLysAlaLeuGlyGlu-----CysValValIleThr 304
QY 2504 CTCAGTGTAAAGTGCAG 2521
Db 305 LeuGluGluMetLeuArg 310

RESULT 10
AB3440
acetylsermidine deacetylase (EC 3.5.1.48) [imported] - Brucella melitensis (strain 1
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AB3440
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD3252; PMID:11756688
A:Accession: AB3440
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <KUR>

```

A:Cross-references: GB:AE008917; PIDN:AAL52685.1; PID:gl17983511; GSPDB:GN00190
A:Experimental source: strain 16M

C:Genetics:
A:Gene: BMEI1504
A:Map position: I
C:Keywords: hydrolase

Alignment Scores:

Pred. No.: 1,23e-15 Length: 337
Score: 438.50 Matches: 116
Percent Similarity: 47.84% Conservative: 50
Best Local Similarity: 33.43% Mismatches: 142
Query Match: 8.20% Indels: 39
DB: 2 Gaps: 11

US-09-502-945-2 (1-2885) x AB3440 (1-337)

```
QY 1481 ACCACAGTGTGCTACGACAGTTCATGCTAAAGCACCGAGTCATGTGCGGGACACA 1540
Db 23 ThrArgLeuTyrTrpHisProIleTyrLeuGluHisLeuThrProGly----- 40
QY 1541 CAGGTGCACCTCAGCATGTGCGCGATCCAGAGATCTGGTCCCGGTGCGAGGAGACA 1600
Db 41 -----HisProGluArgProAspArgIleArgAlaLeuMetSerGluLeuGluGlyPro 58
QY 1601 GGCTGTGTAGCAAGTGCAGCGGATCCGAGGTGCGCAAGCCAGCTAGATGAGATCCAG 1660
Db 59 AspPheTyr---ArgLeuAspArgValGluAlaProHisAlaGlyGluAlaIleLeu 77
QY 1661 ACAGTGCAGTCTCAATACACACACCTGCTGTATGGGACCATGCTCCCTCAACCGGAGAAG 1720
Db 78 LeuAlaHisProGluGluHisLeu-----GluAlaValArgSerLys 91
QY 1721 CTAGACAGCAAGAAGTGTCTGCTGCCATCAGCCAGAGATGATGCTGTGCTGCTTGT 1780
Db 92 IleProGluProValGluAspGlyGluAlaSerGlnProIle----- 105
QY 1781 GGGGGCATGGGTGACAGTGCACCGGTGGAATGAGATGCATCTCCAGTGGCTGTG 1840
Db 106 -----ValLysLeuAspGlyAspThr---TyrValSerProLysSerMetAspAlaAla 122
QY 1841 CGCATGCGAGTGGCTGCTGCTGGAGTGGCTTCAAGTGGCTGCGAGAGAGCTCAAG 1900
Db 123 LeuThrAlaIleGlyAlaAlaMetAlaValAspValMetSerGlyAlaAlaAsp 142
QY 1901 AATGATTTGCCATCATCCGCCCCCAGGACACCGCGAGGAATCCACAGCCATGGGA 1960
Db 143 AsnValPheValAlaSerArgProProGlyHisHisAlaGluArgSerArgAlaMetGly 162
QY 1961 TTCTGCTTCTTCAACTCTGTAGCCATCACCGCAAACTCTTACAGCAGAAAGTTGAACGTG 2020
Db 163 PheCysValPheAsnAsnIleAlaIleAlaAlaArgHisAlaGlnArgHisGlyLeu 182
QY 2021 GCGAAGTCTCATCTGAGTGGGACATCCACATGGCAATGGCACCCAGAGCGGTTTC 2080
Db 183 GluArgIleAlaIleValAspTrpAspValHisHisGlyAsnGlyThrGlnAspIlePhe 202
QY 2081 TACAATGACCCCTCTGCTGTACATCTCTGTCATCGCTATGACACACCGGAACCTCTTT 2140
Db 203 LysAspAspProGlyValMetPheCysSerThrHisGlnPhePro-----LeuTyr 219
QY 2141 CCAGGCTCTGGGCTCCTCGAAGAGGTTGTTGGAGGACCGGCTGGGTGACATGTGAAC 2200
Db 220 ProGlySerGlyAspLysHisGluThr-----GlyValGlyAsnIleValAsn 235
QY 2201 GTGGCATGACAGGAGGTGTGACCCGCCATTTGGAGAGCTGGAGTACCTTACAGCCTTC 2260
Db 236 AlaPro-----LeuSerProAsnThrGlySerArgGluPheArgGluAlaPhe 251
QY 2261 AGGACAGTGGTATGCCCATTTGCCACAGATTTCTACCTGATGTGGTCTAGTCTCCGCC 2320
Db 252 AsnSerArgIleLeuProAlaLeuAspAsnPheArgProAspLeuIleLeuSerAla 271
```

```
QY 2321 GGCTTTGATGCTGTTGAAGGACATCTG---TCTCCTCTGGGTGGCTACTCTGTCAACGCC 2377
Db 272 GlyPheAspAla-----HisPheArgAspProLeuAlaGluIleAsnLeuAspGlu 288
QY 2378 AGATGTTTGGCCACTTGACAGGAGCTGATGACCTCGCA-----GGGGCC 2425
Db 289 SerAspPheAspTrpAlaThrGlyLysLeuMetGluArgAlaGluArgPheCysAspHis 308
QY 2426 CCGGTGCTGCTGCCCTCGAGGAGGCCATGACTTACCGCCATCTGTGATCCCTCTGAA 2485
Db 309 ArgLeuValSerValLeuGluGlyGlyTyrAspLeuGluGlyLeuSerGlnSerAlaSer 328
QY 2486 GCTTGTGTCTCGGCTCTGCTC 2506
Db 329 ValHisIleThrArgLeuLeu 335
```

RESULT 11

T05998
hypothetical protein F17M5_230 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C:Accession: T05998
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15263
A:Accession: T05998
A:Molecule type: DNA
A:Residues: 1-359 <BEY>
A:Cross-references: EMBL:AL035678; GSPDB:GN00062; ATSP:F17M5_230
A:Experimental source: cultivar Columbia; BAC clone F17M5
C:Genetics:
A:Gene: ATSP:F17M5_230
A:Map position: 4
A:Introns: 12/3; 37/3; 71/3; 103/1; 120/2; 199/3; 260/2
C:Superfamily: acetylcholine aminohydrolase; RPD3/acuC homology

Alignment Scores:

Pred. No.: 2,55e-15 Length: 359
Score: 432.50 Matches: 108
Percent Similarity: 50.45% Conservative: 59
Best Local Similarity: 32.63% Mismatches: 131
Query Match: 8.09% Indels: 33
DB: 2 Gaps: 9

US-09-502-945-2 (1-2885) x T05998 (1-359)

```
QY 1547 CACCTGAGCATGTGCGCGATCCAGAGCATCTGTCTCCGCTGCAGGAGAGCCCTG 1606
Db 15 HisProGluCysSerAlaArgValProAlaIleValAsnAlaLeuGluMetAsnGluLeu 34
QY 1607 CTAGCAAG-----TGCAGCGGATCCGAGGTGCGAAAGCCACGCCTA 1648
Db 35 ThrProLysPheArgGlySerGlnIleLeuGluAlaAsnPheLysThrAlaThrVal 54
QY 1649 GATGAGATCCAGACAGATGCTGATACACACCTGCTCTATGGACACGCTCCCTC 1708
Db 55 GluAspIleAlaAsnValHisAspLysAlaTyr-----ValPheGlyLeuGluLysAla 72
QY 1709 AACCGGAGAGCTAGACAGCAAGATTTGCTCGCTCCATCAGCAGAGATGTATGCT 1768
Db 73 MetAspGluAlaSerAspSerGlyLeuIlePheIleGlySerGlyProThrTyrAla 92
QY 1769 GTGCTGCTTGTGGGGCATCGGGGTGGACACTGACACCGTGTGAATGAGATGCTCC 1828
Db 93 -----ThrSerThrValThrAlaLeuCysSerSer 102
QY 1829 TCCAGTGTGTCGCATGCGAGTGGCTGCTGTGAGAGTGGCTTCAAGTGGCTGCA 1888
Db 103 AlaGlyAlaGlyMetAlaLeuValAspSerValValMetSerArgLeuThrIleAsnSer 122
QY 1889 GGAGAGCTCAAGAATGGATTTGCCATCATCGGCCCCCAGGACACCGCGGAGGAATCC 1948
Db 123 ValAspProProIleGlyPheAlaLeuIleArgProProGlyHisHisAlaValProLys 142
```


[illegible]

Db	20	HisProGluArgProAspArgLeuArgSerLeuAsnIleAlaLeuGluHisProAsn---	38
QY	1607	CTTAGCAAGTGGAGCGGATCCGAGGTGCGAAAGCCACGCTAGATGAGATCGACAGAGTG	1666
Db	39	PheGluArgLeuAspArgGlySerGluAlaProGlnAlaAsnGluAspAlaValLeuLeuAla	58
QY	1667	CACCTTGAATACCAACCCCTGCTTATGGGACCAGTCCCTCAACCGCGAGAGCTAGAC	1726
Db	59	HisProGluGluHisLeuLeu-----AlaValMetArgGlnValProGlu	73
QY	1727	AGCAAGAAGTTCTCGGTCCCATCAGCCAGAAGATGATGCTGTGCTGCTTGTGGGGC	1786
Db	74	GluAsp-----GlyIleIleasnArg-----	80
QY	1787	ATCGGGGTGGAGCATGACACCGCTTGGAAATGAGATGCACCTCCCTCCAGTGCCTGCCGATG	1846
Db	81	-----IleGluSerAspThr---TyrLeuSerProLysSerLeuGlnAlaLeuThr	97
QY	1847	GCAGTGGGTGCTGCTGGAGCTGGCCCTTCAAGTGGCTGCAGGAGAGCTCAGAATGGA	1906
Db	98	GlyIleGlyAlaAlaMetAlaAlaValAspValPheSerGlyAlaAlaAspAsnVal	117
QY	1907	TTTGGCCATCATCGGCCCCAGGACACACCGAGGAATCCACAGCCATGGGATCTGC	1966
Db	118	PheValAlaAlaArgProGlyHisAlaGluThrAlaLysAlaMetGlyPheCys	137
QY	1967	TTCTTCAACTCTGTAGCATCAACCACAAATCTCTACAGAGAAAGTTGAACGTGGCGAAC	2026
Db	138	LeuPheAsnAsnValAlaIleAlaAlaArgHisAlaGlnLysAlaHisGlyAlaGluArg	157
QY	2027	GTCTCATCGTGGACATTCACCATGGCAATGCCATGCCACCCAGCAGGGTGTCTACAAT	2086
Db	158	IleAlaIleIleAspValHisHisGlyAsnGlyThrGlnAspIlePheTrpAsn	177
QY	2087	GACCCCTCTGTGCTACATCTCTCTGCTGCTATGACAACGGGAAGTCTTCTTCCAGGC	2146
Db	178	AspIleSerValLeuPheCysSerThrHisGlnMetPro-----LeuTyrProTrp	194
QY	2147	TCTGGGGTCTCTGAAGAGTTGGTGAGGACCGAGCGTGGGGTACAATGAAAGTGCGCA	2206
Db	195	SerGlyAspLysAsnGluThr-----GlyValLysAsnAsnIleValAsnAla	210
QY	2207	TGGACAGAGGTGTGGACCCGCCCATTTGGAGAGCTGGAGTACCTTACAGCCTTCAGGACA	2266
Db	211	-----ProLeuSerProAsnThrGlySerAspHisPheArgGluAlaPheLysSer	227
QY	2267	GTGTTGATGCCCATTCGCCACAGTCTCACCTGATGGTCTAGTCTCGCGCGGGTTT	2326
Db	228	ArgValLeuProAlaIleAlaAspPheSerProAspLeuIleLeuIleSerAlaGlyPhe	247
QY	2327	GATGCTGTTGAAGGACATGTCTCTCTGGTGGCTACTCTGCACCGCCAGATGTTTT	2386
Db	248	AspAla-----HisHisArgAspProLeuAlaGlnIleAsnLeuValGlyGluAspPhe	265
QY	2387	GGCCACTTGTACAGGACGTGTATGCCCTG-----GCAGGGGCGCGGTGGTG	2434
Db	266	AspTrpAlaThrGlyArgLeuLeuGluMetAlaAspLysTyrAlaSerAsnArgVal	285
QY	2435	CTGGCCCTGGAGGGCCATGACTTGACCGCCATCTGTATGCTCT	2482
Db	286	SerLeuLeuGluGlyGlyTyrAspLeuGlyLeuAlaGluSerAla	301

RESULT 13

AF2668

deacetylase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002

C:Accession: AF2668

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chera, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, Y.; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

Alignment scores:	
Pred. No.:	1.14e-14
Score:	420.00
Length:	311
Matches:	107
Conservative:	50
Percent Similarity:	49.68%
Mismatches:	121
Best Local Similarity:	33.86%
Query Match:	7.86%
Indels:	38
Gaps:	10
DB:	2

US-09-502-945-2 (1-2885) x D97450 (1-311)

QY 1547 CACCCTGAGCATGCTGGCCCGGATCCAGAGCATCTGGTCCCGGCTGCAGGAGACAGGCCCTG 1606

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.

A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AF2668

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-311 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL41764.1; PID:g17739116; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu0748

A:Map position: circular chromosome

C:Superfamily: acetylpolylamine aminohydrolase; RPD3/acuc homology

Alignment Scores:

Pred. No.:	1.14e-14	Length:	311
Score:	420.00	Matches:	107
Percent Similarity:	49.68%	Conservative:	50
Best Local Similarity:	33.86%	Mismatches:	121
Query Match:	7.86%	Indels:	38
DB:	2	Gaps:	10

US-09-502-945-2 (1-2885) x AF2668 (1-311)

QY 1547 CACCTGAGCATGTCGGCGGATCCAGAGCATCTGTCGCCGCTGCAGAGACAGAGCGCTG 1606

Db 20 HisProGluArgProAspArgLeuArgSerLeuAsnIleAlaLeuGluHisProAsn--- 38

QY 1607 CTTAGCAAGTCGGAGCGATCCGAGGTCGCAAGCCAGCTAGATGATCCAGACAGT 1666

Db 39 PheGluArgLeuAspArgLysGluAlaProGlnAlaAsnGluAspAlaValLeuAla 58

QY 1667 CACTCTGAATACACACCCCTCTATGAGGACCATGTCCTCAACCGGACAGAGCTAGAC 1726

Db 59 HisProGluHisLeuLeu-----AlaValMetArgGlnValProGlu 73

QY 1727 AGCAAGAAGTTGCTGCTGCCATCAGCAGCAAGATGATGCTGCTGCTGCTGCTGCTG 1786

Db 74 GluAsp-----GlyGluIleAsnArg----- 80

QY 1787 ATCGGGGTGACAGTACACCGTGTGGATGATGATGATGATGATGATGATGATGATG 1846

Db 81 -----IleGluSerAspThr-----TyrLeuSerProLysSerLeuGlnAlaAlaLeuThr 97

QY 1847 GCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1906

Db 98 GlyIleGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 117

QY 1907 TTTGCCATCATCGGCGCCCGGAGGACCCAGCGGAGGAAATCCACAGCATGGGATTTCTGC 1966

Db 118 PheValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 137

QY 1967 TTTCTCACTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2026

Db 138 LeuPheAsnAsnValAlaIleAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 157

QY 2027 GTCTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2086

Db 158 IleAlaIleIleAspThrAspValHisHisGlyAsnGlyThrGlnAspIlePheThrpsn 177

QY 2087 GACCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2146

Db 178 AspIleSerValLeuPheCysSerThrHisGlnMetPro-----LeuTyrProThr 194

QY 2147 TCTGGGCTCTCCTGAAGAGTTGCTGGAGGACCAAGCGGTGGGTGTACATGTGAACGTG 2206

Db 195 SerGlyAspLysAsnGluThr-----GlyValLysAsnAsnIleValAsnAla 210

QY 2207 TGGACAGGAGGTGTGACCCCGCCCATTTGGAGAGCTGGAGTACCTTACAGCTTACAG 2266

Db 211 -----ProLeuSerProAsnThrGlySerAspHisPheArgGluAlaPheLysSer 227

QY 2267 GTGGTGTATGCCCATTTGCCACGAGTTCTACCTGATGTGCTCTAGTCTCCGCGGGTTT 2326

Db 228 ArgValLeuProAlaIleAlaAspPheSerProAspLeuIleLeuIleSerAlaGlyPhe 247

QY 2327 GATGCTGTTGAAGGACATCTGCTCTCTGCTGGTGGTACTCTCTCACCAGCATGTTT 2386

Db 248 AspAla-----HisHisArgAspProLeuAlaGlnIleAsnLeuValGlyGluAspPhe 265

QY 2387 GGCACCTGACAGGAGGAGCTGATGACCTG-----GCAGGGGCGCGGGTGGT 2434

Db 266 AspTrpAlaThrGlyArgLeuLeuGluMetAlaAspLysTyrAlaSerAsnArgValVal 285

QY 2435 CTGGCCCTGAGGAGGAGGAGCTGACCTGACCGCCCTGCTGCTGCTGCTGCTGCT 2482

Db 286 SerLeuLeuGluGlyGlyArgLeuGluGlyLeuAlaGluSerAla 301

RESULT 14

T27101

hypothetical protein Y51H1A.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27101

R:Smyle, R.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20309

A:Accession: T27101

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-517 <WIL>

A:Cross-references: EMBL:AL032644; PIDN:CAA21669.2; GSPDB:GN00020; CESP:Y51H1A.5

A:Experimental source: clone Y51H1A

C:Genetics:

A:Gene: CESP:Y51H1A.5

A:Map position: 2

A:Introns: 41/3; 84/3; 218/1; 408/1

Alignment Scores:

Pred. No.:	5.98e-14	Length:	517
Score:	406.50	Matches:	127
Percent Similarity:	47.48%	Conservative:	71
Best Local Similarity:	30.46%	Mismatches:	152
Query Match:	7.60%	Indels:	67
DB:	2	Gaps:	17

US-09-502-945-2 (1-2885) x T27101 (1-517)

QY 1490 GTGGTCTACGACACCTTCATGCTAAAGACACAGTCATGTCGGGGAACACACACAGTCGAC 1549

Db 9 ValIlePheAspHisGlnGluArgHisAspGlnProTrpProSerTyrHisIle--- 27

QY 1550 CTGAGCATGTCGGCGGATCCAGAGCATCTGTCGCCGCTGCAGGACAGGCGCTGCTT 1609

Db 28 ---GluValProArgArgLeuAspAlaIleLeuGluArgLeuAsnThrThrLysLeuLeu 46

QY 1610 AGC-----AAGTGGCGCGGATCCGAGTCCCAAGCCACCTAGATGAGATCCAGACA 1663

Db 47 ThrAspProArgIleGluHisIleProArgGluAlaGluGluSerGluIleLeuAla 66

QY 1664 GTGCACTCTGAATACACACCTGCTCTATGGACCATGTCCTCAACCGGCGGAGAAGTA 1723

Db 67 ValHisThrLysArgTyrVal-----AspAspVal 76

QY 1724 GACACAGAAGATGCTCGTCCCATCAGCCAGAGATGATGTGCTGCTGCTGCTGCTGCTG 1783

Db 77 LysSerThrGluThrMetThrValGluGlnGluSerPheCys----- 91

QY 1784 GCATCGGGGTGCAGAGTGCACACCTGCGGAATGAGATGACCTCTCCCTCCAGCTGCT 1837

Db 92 -----ThrLysTyrGluAspIleTyrValAsnSerAlaThrTrp 104

QY 1838 -----GTGCGCATGGCAGTGGGCTGCTGCTGGAGCTGGCCTTCAAGCTGCTGCTG 1891

Db 105 HisArgAlaLysLeuAlaAlaGlyAlaSerIleAspLeuMetThrSerValMetAlaAla 124

QY 1892 GAGCTCAGAAATGGATTGGCCATCATCCGCGCCCGCCAGACACACCGCCGCGGAATCCACA 1951
Db 125 Lys---ArgProGlyIleAlaPheIleArgProProGlyHisAlaMetProAspGlu 143
QY 1952 GCCATGGGATTCCTCTTCAACTCTTAGCCATCACCGCAAACTCCTACAGCAGAAG 2011
Db 144 GlyCysGlyPheCysIlePheAsnAsnValAlaIleAlaAlaLysAlaIleGln--- 162
QY 2012 TTGAACGTGGGCAAGGTCTCATCTGACCTGGGACATTCACCATGGCAATGGCACCCAG 2071
Db 163 ---AsnGlyGlnLysValLeuIleValAspTyrAspValHisAlaGlyAsnGlyThrGln 181
QY 2072 CAGGCGTTC-----TACAATGACCCCTCTGCTCTACATCTCTCATCTGCTATGAC 2125
Db 182 GluCysValGlnMetGlyGlnGlyAsnValGlnLeuIleSerIleHisArgTyrGln 201
QY 2126 AACGGGAAGTCTTTCCAGGCTCTCGGGCTCCTGAAGAGTGTGGTGGAGGACCGCGT 2185
Db 202 AsnGlyHisPheTyrPro-----AsnMetProGln-----ThrGlyIle 214
QY 2186 GGGTACATGTGAACGTGGCATGGACAGGAGGTGGGACCCGCCCATTTGGA----- 2236
Db 215 TyrHisAsn-----TyrLysAsnThrIleAsnLeuProLeuAsnThrIleGly 230
QY 2237 -----GACGTGGAGTACCTTACAGCCTTCAGGACAGTGGTATGCCATTGCCACGAG 2290
Db 231 LeuThrAspAlaAspTyrHisAlaLeuPheThrHisIleIleLeuProThrIleHisAla 250
QY 2291 TTCTACCTGATGTCTAGTCTCCGCGGGTTTGTGTTGTTGAAGGACATCTGTCT 2350
Db 251 PheGlnProAspLeuLeuValSerSerGlyPheAspAlaSerIleGly-----Asp 268
QY 2351 CCTCTGGGTGGTACTCTCTACCCGCCAGATGTTTGGCCATTGACAGGAGGTGATG 2410
Db 269 ProGluGlySerMetGlnValThrProAlaGlyPheAlaThrMetIleArgMetLeuIle 288
QY 2411 ACCCTGGCAGGGGGCGGTGTGTGGCCCTGGAGGAGGACATGACATGACCGCCATC 2470
Db 289 Asp---ThrGlyIleProValAlaAlaLeuLeuGluGlyGlyTyrPheLeuAspAlaLeu 307
QY 2471 TGTGATGCTTGAAGCTGTGTCTCGGCTCTGCTCAGT-----GTAAAG 2515
Db 308 AlaAlaAspSerGluTrpValLeuArgAlaLeuLeuGlyGluGluPheProArgIleArg 327
QY 2516 CTGCACCCCTGGATGAGCAGCTCTTGACGAAACCCCAACATCAACGACGTGGCCAG 2575
Db 328 ValGluLysIleHisSerAlaIle-----AlaAspThrIleGlyArg 341
QY 2576 CTAGAGAAAGTATCATCAGATCCAGAGCAACACTGGAGTGTGTGAGAAAGTTCGCGCT 2635
Db 342 ValValLysArgTyrGluGlySerCysProPhePheLysIleValGlnGluLeuArgGly 361
QY 2636 GGTCTGGGCGGTCTCTGCGAGGGGCCCAAGCAGGTGAGACCGAAGACCC 2686
Db 362 IleLeuGly-----CysArgValAlaGluAspGlnGlnGluAla 376

RESULT 15

D83174
probable acetylpolymine aminohydrolase PA3774 [imported] - Pseudomonas aeruginosa (stra
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83174
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoquchi, S.D.; Warriner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: D83174
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-380 <STO>

A:Cross-references: GB:AE004796; GB:AE004091; NID:g9949939; PIDN:AG07161.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3774

Alignment Scores: 4.84e-13 Length: 380
Pred. No.: 389.00 Matches: 102
Score: 389.00
Percent Similarity: 46.72% Conservative: 62
Best Local Similarity: 29.06% Mismatches: 129
Query Match: 7.28% Indels: 58
DB: 2 Gaps: 9

US-09-502-945-2 (1-3885) x D83174 (1-380)

QY 1550 CCTGAGCATGCTGCCCGATCCAGAGCATCTGGTCCCGGTGCAGGAGAGCGCTGCTT 1609
Db 43 ProGluThrLysArgArgLeuLysSerLeu-----LeuAspValSerGlyLeuThr 59
QY 1610 AGCAAGTGGCAGCGGATCCGAGGTCCGAAAGCCACGCTAGATGAGATCCAGACAGTGCAC 1669
Db 60 AlaArgLeuGlnLeuArgSerAlaProProAlaSerAspGluAspLeuLeuArgValHis 79
QY 1670 TCTGAATACACACCCCTGCTCTATGGGACCATGCTCCCTCAACCGGACAGCTAGACAGC 1729
Db 80 ProAlaHisTyrLeu----- 84
QY 1730 AAGAAGTGTCTCGTCCCATCATCAGCAGAGATGATGCTGTGCTGCTTGTGGGGCATC 1789
Db 85 -----GluArgPheLysAlaLeuSerAspAlaGlyGlyGly 96
QY 1790 GGGTGGACAGTACACCGCTGTGGAATGAGATGACCTCCTCCAGTGTGTGCCGATGGCA 1849
Db 97 SerLeuGlyGlnAspAlaProIleGlyPro---GlySerTyrGluIleAlaArgLeuSer 115
QY 1850 GTGGGTCCTGCTGAGCTGCGCTTCAAGTGTGCTGCAGGAGAGCTCAAGATGGATTT 1909
Db 116 AlaGlyLeuAlaIleAlaAlaLeuAspAlaValLeuAlaGlyGluAlaAspAsnAlaTyr 135
QY 1910 GCCATPACCCGCCCCAGGACACACCCGAGGAATCCACAGCCATGGGATTTCTGTTTC 1969
Db 136 SerLeuSerArgProProGlyHisHisCysLeuProAspGlnAlaMetGlyPheCysPhe 155
QY 1970 TTCAACTCTAGCCATCACCGCAAACTCTACAGCAAGATTTGAAGTGGGCAAGTGC 2029
Db 156 PheAlaAsnIleAlaValAlaIleGluAlaAlaLysAlaArgHisGlyValGluArgVal 175
QY 2030 CTCATCTGGGACTGGGACATTCACCATGGCAATGGCACCACCGGCGCTTCTACATGAC 2089
Db 176 AlaValLeuAspTrpAspValHisHisGlyAsnGlyThrGlnAlaIleTyrTyrArgArg 195
QY 2090 CCTCTGTGCTCTACATCTCTGCTGCTATGACAAACGGGAACCTTTTCCAGGCTCT 2149
Db 196 AspAspValLeuSerIleSerLeuHisGln-----AspGlyCysPheProProGlyTyr 213
QY 2150 GGGCTCTCTGAAGAGTGTGGTGGAGCACCGCTGGGGTACAAATGTGAACGTGGCATGG 2209
Db 214 SerGlyAlaGluAspIleGlyGluAspArgGlyArgGlyPheAsnLeuAsnValPro--- 232
QY 2210 ACAGAGGTGTGGACCCCGCCCATTTGGAGACCTGGAGTACCTTACAGCCCTCAGGACATG 2269
Db 233 -----LeuLeuProGlyGlyGlyHisAspAlaTyrMetGlnAlaMetGlnArgIle 249
QY 2270 GTGATGCCCATTTGCCACGAGTTCTCAGCTGATGGTCTCTAGTCTCCGCGGTTTTCAT 2329
Db 250 ValLeuProAlaLeuGluArgPheArgProGlnLeuIleValAlaSerGlyPheAsp 269
QY 2330 GCTGTGTGAAGACATCTCTCTCTGGTGGCTACTCTGTCCACCGCCAGATGTTTGGC 2389
Db 270 Ala-----AsnAlaValAspProLeuAlaArgMetGlnLeuHisSerAspSerPheArg 287
QY 2390 CACTTGACCGACGCTGATGACCTG-----GCAGGGGCGCGGTGGTGTG 2437
Db 2390 CACTTGACCGACGCTGATGACCTG-----GCAGGGGCGCGGTGGTGTG 2437

Db 288 AlaMetThrIaMetValArgAspAlaAlaGluArgHisAlaGlyGlyArgLeuValVal 307
QY 2438 GCCCTGGAGGAGGCCCATGACTTGACCGCCCATCTGTGATGCTCTCTGAAGCTTGTGTC--- 2494
|||||
Db 308 ValHisGluGlyGlyTyr-----SerGluAlaTyrValPro 319
|||||
QY 2495 -----TCGGCTCTGCTCAGTGTAAAGCTGCAGCCC 2524
|||
Db 320 PheCysGlyLeuAlaValIleGluGluLeuSerGlyValArgSerAlaValArgAspPro 339
|||
QY 2525 TTGGATGAGGCAGTCTTGCAGCAAAAGCCCAAC 2557
|||
Db 340 LeuArgAspPheIleGluLeuGlnGlnProAsn 350
|||

Search completed: March 21, 2003, 13:04:54
Job time : 78.3908 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 13:00:02 ; Search time 30.8914 Seconds

(without alignments)
9985.861 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346

Sequence: 1 ggaattctctgtcgaagt.....aaacaaaagttaaaattt 2885.

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 442306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09502945/runat_14032003_101101_19301/app_query.fasta_1.10979
-DB=PublishedApplications-AA-QFMT-fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09502945 @CGN_1.1.60 @runat_14032003_101101_19301
-NCPU=6 -ICPU=3 -NO_XIPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4793	89.7	1122	9	US-10-072-094-8
2	4793	89.7	1122	9	US-10-072-094-91
3	3612.5	67.6	717	10	US-09-817-913-9
4	3612.5	67.6	717	10	US-09-817-538-9

5	2691	50.3	1084	9	US-10-072-094-7
6	2685	50.2	967	10	US-09-817-913-7
7	2685	50.2	967	10	US-09-817-538-7
8	2688	49.9	1069	9	US-10-072-094-87
9	2579	48.2	1011	9	US-10-072-094-89
10	2212.5	41.4	780	9	US-10-072-094-93
11	2184.5	40.9	1141	9	US-10-072-094-107
12	1988	37.2	879	9	US-10-072-094-90
13	1674.5	31.3	855	9	US-10-072-094-9
14	1671.5	31.3	855	10	US-09-817-913-13
15	1671.5	31.3	855	10	US-09-817-538-13
16	961.5	18.0	590	9	US-10-072-094-99
17	893.5	16.7	495	9	US-10-072-094-95
18	771.5	14.4	1215	10	US-09-976-280A-4
19	771.5	14.4	1215	10	US-09-976-280A-22
20	766.5	14.3	1215	10	US-09-817-913-11
21	766.5	14.3	1215	10	US-09-817-538-11
22	723	13.5	163	9	US-10-072-094-4
23	705.5	13.2	687	9	US-09-976-280A-3
24	672	12.6	706	9	US-10-072-094-6
25	623.5	11.7	673	10	US-09-976-280A-2
26	595	11.1	342	9	US-09-911-150-4
27	543	10.2	175	9	US-10-072-094-5
28	445	8.3	112	9	US-10-072-094-21
29	413	7.7	105	9	US-10-072-094-2
30	374.5	7.0	335	10	US-09-971-309-80
31	316.5	5.9	480	10	US-09-347-331-9
32	306.5	5.7	488	10	US-09-925-300-1563
33	296.5	5.5	150	9	US-09-911-150-6
34	291	5.4	481	10	US-09-817-913-1
35	291	5.4	481	10	US-09-817-538-1
36	289	5.4	576	10	US-09-347-331-11
37	284.5	5.3	2530	9	US-10-184-644-155
38	279.5	5.2	377	10	US-09-784-877-2
39	279.5	5.2	377	10	US-09-817-913-15
40	279.5	5.2	377	10	US-09-817-538-15
41	278.5	5.2	2457	9	US-10-184-644-569
42	278	5.2	517	10	US-09-883-720-8
43	277.5	5.2	458	10	US-09-883-720-2
44	277.5	5.1	1903	9	US-10-184-644-455
45	277	5.1	1815	9	US-10-184-644-19

ALIGNMENTS

RESULT 1
US-10-072-094-8
; Sequence 8, Application US/10072094
; Publication No. US20030025538A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, DONALD
; APPLICANT: LORENZI, MATTHEW
; APPLICANT: ATTAR, RICARDO
; APPLICANT: GOTTARDIS, MARCO
; TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
; FILE REFERENCE: 3053-4145U1
; CURRENT APPLICATION NUMBER: US/10/072, 094
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,296
; NUMBER OF SEQ IDS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-094-8

Alignment Scores:
Pred. No.: 7.84e-228 Length: 1122
Score: 4793.00 Matches: 927
Percent Similarity: 99.57% Conservative: 3
Best Local Similarity: 99.25% Mismatches: 4

Query Match: 89.66% Indels: 1
DB: 9 Gaps: 0

US-09-502-945-2 (1-2885) x US-10-072-094-8 (1-1122)

```
QY 2 GAATTCCTCTTGTCAAGTCAAGAGCCACACAGCGGCTCAACCATTCCTCCCA 61
DB 189 GluPheLeuLeuSerLysSerLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 208
QY 62 CAGCACCCCAATGCTGGGAGCCACCATGCTCTTGGACACAGTTCCTCCCGAG 121
DB 209 GlnHisProLysCysTrpGlyAlaHisAlaSerLeuAspGlnSerSerProGln 228
QY 122 AGCGGCCCCCTGGGACGCTCCCTCCACAACTGCCTTTCCCTGGGCCCTACACAGT 181
DB 229 SerGlyProGlyThrProProSerTyrLysLeuProLeuProGlyProTyrAspSer 248
QY 182 CGAGAGGACTTCCCTCCGGAACACAGCCTCTGAACCCAACTTGAAGTGGCTCAAGG 241
DB 249 ArgAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg 268
QY 242 TAAACACAGAAGTGGCTCAGCGGAGAGCAGTCCCTCTCGTCGCAAGGATGGACT 301
DB 269 LeuLysGlnLysValAlaGluArgSerSerProLeuLeuArgLysAspGlyThr 288
QY 302 GTATTAGCACCTTTAAGAGAGAGCTGTTGAGATCACAGGTGCCGGGCTCGGCGCTCG 361
DB 289 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyAlaGlyProGlyAlaSer 308
QY 362 TCGGTGTAAACAGCGCCCGGCTCCGGCCCGAGCTCTCCCAACAGCTCCACAGCACC 421
DB 309 SerValCysAsnSerAlaProGlySerGlyProSerSerProAsnSerSerHisSerThr 328
QY 422 ATCGCTGAGAATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCAG 481
DB 329 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGln 348
QY 482 CACCGAGCCCTCCCTCTGGACAGCTCCCAACCAAGTCTCAGGCTCTACAGCTCCTCTCT 541
DB 349 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTyrThrSerProSer 368
QY 542 CTGCCCCAATCTCCCTAGGCTGACGCGCAGCGTCACTGTGTCACCAACTCACACCTCACT 601
DB 369 LeuProAsnIleSerLeuGlyLeuGlnAlaThrValThrValThrAsnSerHisLeuThr 388
QY 602 GCCTCCCCAGCTGTCGACACAGAGAGGCGCGAGACAGCGCCCTCCAGTCCCTCGCG 661
DB 389 AlaSerProLysLeuSerThrGlnGlnGluAlaGluArgGlnAlaLeuGlnSerLeuArg 408
QY 662 CAGGTTGGCAGCTGACCGGCAAGTTTCATGAGCACATCTCTATTCTGGCTGCCCTGCTG 721
DB 409 GlnGlyGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCysLeuLeu 428
QY 722 GCGGTGGCACTGGAGGGGACGGGACCCCGCCAGCGCATGCCCTCCCTGTGTCAGCATG 781
DB 429 GlyValAlaLeuGluGlyAspGlySerProHisGlyHisAlaSerLeuLeuGlnHisVal 448
QY 782 CTGTTGCTGGACAGCCCGGACGACAGACACCTCATTTGCTGTGCACTCCACGGGACG 841
DB 449 LeuLeuLeuGluGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 468
QY 842 TCCCCACTAGTACGGGTGAACCTGTGGCCACCAAGATCGGAGCTAGGCAAGCTCCCG 901
DB 469 SerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLysLeuPro 488
QY 902 CGGCATCGGCCCTGAGCGCACTCACTAGTCTCACCCTGCGGACAGAGTCCCCAGGCCCTG 961
DB 489 ArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGlnAlaLeu 508
QY 962 CAGCAGCTGTGTATGAACAACAGCACCAGCAGCTTCTCTGGAGAAGCAGAGCAGCAG 1021
DB 509 GlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnLysGlnGlnGln 528
```

```
QY 1022 CTACAGCTGGSCAAGATCCTCACCAAGACAGAGGGAGTGCCCGAGGCACCCACCCAC 1081
DB 529 LeuGlnLeuGlyLysIleLeuThrLysThrGlyGluLeuProArgGlnProThrHis 548
QY 1082 CCTGAGAGACAGAGGAGGAGCTACCGAGCAGCAGAGGTCTTGCTGGGGGGAGACC 1141
DB 549 ProGluGluThrGluGluLeuThrGluGlnGlnGluValLeuLeuGlyGluGlyAla 568
QY 1142 CTGACCATTGCCCGGAGGCTCCACAGAGAGTGGAGAGCAGACAGAGAGCTCGAGGAG 1201
DB 569 LeuThrMetProArgGluGlySerThrGluSerGluSerThrGlnGluAspLeuGlu 588
QY 1202 GAGCAGCAGGAGGAGTGGGAGGAGGAGGAGGATTGCATCCAGGTTAAGGACGAGAG 1261
DB 589 GluAspGluGluGluAspGlyGluGluAspCysIleGlnValLysAspGluGlu 608
QY 1262 GGCAGAGTGTGTGTGAGGAGGCGCCGACTTGAGAGAGCTGTGTGGATACAAAAA 1321
DB 609 GlyGluSerGlyAlaGluGluGluProAspLeuGluGluProGlyAlaGlyTyrLys 628
QY 1322 CTGTTCTCAGATGCCCAACCGCTCAACCTTTGAGGTGTACCAAGCGCCCTCAGGCTG 1381
DB 629 LeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyrGlnAlaProLeuSerLeu 648
QY 1382 GCCACTGTGCCCCACCAAGCCCTGGGCGGTACCCAAATCCTCCCTGCTGCCCTGGGGGC 1441
DB 649 AlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaProGly 668
QY 1442 ATGAAGAACCCCGACAGCAACCGCTCAAGCACCTTTCACCAACAGTGTGGTCTACGAC 1501
DB 669 MetLysSerProProAspGlnProValLysHisLeuPheThrThrGlyValValTyrAsp 688
QY 1502 ACGTTTCATGCTAAAGCACACAGTGTGCGGGACACACACAGTGCACCCCTGAGCATGCT 1561
DB 689 ThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGluHisAla 708
QY 1562 GCGCGGATCCAGAGCATCTGTGTCGGGCTGAGGAGCAGGAGCTGCTTAGCAAGTGCAG 1621
DB 709 GlyArgIleGlnSerIleTrpSerArgLeuGlnGluThrGlyLeuLeuSerLysCysGlu 728
QY 1622 CGGATCCGAGTGCAGAACCCAGCTAGATGAGATCCAGAGTGCATCTCTGAATACCAC 1681
DB 729 ArgIleArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGluTyrHis 748
QY 1682 ACCCTGCTCTATGGACCAAGTCCCTCAACCGGAGAGAGTGCAGAGCAAGAGTGTGCTC 1741
DB 749 ThrLeuLeuTyrGlyThrSerProLeuAsnArgGlnLysLeuAspSerLysLysLeuLeu 768
QY 1742 GGTCCCATCAGCAGAGATGTATGCTGTGTCCTTGTGGGGCATCGGGTGGACAGT 1801
DB 769 GlyProIleSerGlnLysMetTyrAlaValLeuProCysGlyGlyIleGlyValAspSer 788
QY 1802 GACACCGTGTGAATGAGATGCACCTCCCTCCAGTGTGTCGTCGATGGCATGGCTGGCTG 1861
DB 789 AspThrValTrpAsnGluMetHisSerSerAlaValArgMetAlaValGlyCysLeu 808
QY 1862 CTGGAGCTGGCTTCAAGTGGCTGCAGGAGAGCTCAAGAATGGATTGGCCATCATCCGG 1921
DB 809 LeuGluLeuAlaPheLysValAlaGlyGluLeuLysAsnGlyPheAlaIleIleArg 828
QY 1922 CCCCCAGGACACACCCCGGAGGAATCCACAGCCATGGGATTTCTTCAACTCTGTA 1981
DB 829 ProProGlyHisHisAlaGluSerThrAlaMetGlyPheCysPheAsnSerVal 848
QY 1982 GCCATCCGCAAACTCCTACAGCAGAGTGAACGTGGGCAAGGCTCTCATCTGGAC 2041
DB 849 AlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeuIleValAsp 868
QY 2042 TGGGACATTCACCATGGCAATGGCACCGAGGCTTCTCAATGACCCCTCTCTGCTC 2101
DB 869 TrpAspIleHisHisGlyAsnGlyThrGlnGlnAlaPheTyrAsnAspSerValLeu 888
QY 2102 TACATCTCTCTCATCGCTATGACAAACGGGAACCTCTTCCAGGCTCTGGGGCTCTCTGAA 2161
```

Db 889 TyrIleSerLeuHisArgTyrAspAsnGlyAsnPhePheProGlySerGlyAlaProGlu 908
Qy 2162 GAGTTGTGTGGAGGACCGGTGGGTACATGTGACATGTGACATGTGACATGTGACATGTG 2221
Db 909 GluValGlyGlyGlyProGlyValGlyTyrAsnValAsnValAlaTrpThrGlyGlyVal 928
Qy 2222 GACCCCCCATTTGGAGAGCTGGAGTACCTTACAGCCTTTCAGACAGTGGTGTGATGCCATT 2281
Db 929 AspProIleGlyAspValGlyTyrLeuThrAlaPheArgThrValValMetProIle 948
Qy 2282 GCCACAGATTCTCACCTGATGTGTCTAGTCTCCGCCGCTTTGATGCTGTGTAAGGA 2341
Db 949 AlaHisGluPheSerProAspValValLeuValSerAlaGlyPheAspAlaValGluGly 968
Qy 2342 CATCTGTCTCTCTGGTGGTACTCTCTACCGCCAGATGTTTGGCCACTTGCACAGG 2401
Db 969 HisLeuSerProLeuGlyGlyTyrSerValThrAlaArgCysPheGlyHisLeuThrArg 988
Qy 2402 CAGCTGATGACCTGGCAGGGGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2461
Db 989 GlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGluGlyGlyHisAspLeu 1008
Qy 2462 ACCGCCATCTGTGATGCTCTGACAGTGTGTCTCGGTCTGCTCAGTGTAAAGCTGCAG 2521
Db 1009 ThrAlaIleCysAspAlaSerGluAlaCysValSerAlaLeuLeuSerValGluLeuGln 1028
Qy 2522 CCCTTGGATGAGGAGCTCTGCACCAAGCCACATCAACGAGTGGTGGTGGTGGTGGTGGTGG 2581
Db 1029 ProLeuSerGluAlaValLeuGlnGlnLysProAsnIleAsnAlaValAlaThrLeuGlu 1048
Qy 2582 AAAGTCATCGAGATCCAGAGCAACACTGGAGCTGTGTGCAGAGTTCGCCGCTGGTCTG 2641
Db 1049 LysValIleGluIleGlnSerLysHisTrpSerCysValGlnLysPheAlaAlaGlyLeu 1068
Qy 2642 GGCGGTCCCTCGCAGGGGGCCACGAGTGTGAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 2700
Db 1069 GlyArgSerLeuArgGluAlaGlnAlaGlyGluThrGluGluAlaGluThrValSerAla 1088
Qy 2701 ATGGCTTGTCTGTGGTGGGGCCGACAGCCCAAGCTGGAGCTGTGTGCAGAGTTCGCCGCTGG 2760
Db 1089 MetalLeuLeuSerValGlyAlaGluGlnAlaGluThrGluGluAlaAlaAlaArgGluHisSer 1108
Qy 2761 CCCAGGGCGGAGGAGGCCCATCGAGCAGGAGGCTGCGCTG 2802
Db 1109 ProArgProAlaGluGluProMetGluGlnGluProAlaLeu 1122

RESULT 2

US-10-072-094-91

; Sequence 91, Application US/10072094
; Publication No. US2003002538A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, DONALD
; APPLICANT: LORENZI, MATTHEW
; APPLICANT: ATTAR, RICARDO
; APPLICANT: GOTTARDIS, MARCO
; TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
; FILE REFERENCE: 3053-4145US1
; CURRENT APPLICATION NUMBER: US/10/072,094
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,296
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 91
; LENGTH: 1122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-094-91

Alignment Scores: 7.84e-228 Length: 1122
Pred. No.: 4793.00 Matches: 927
Score:

Percent Similarity: 99.57% Conservative: 3
Best Local Similarity: 89.25% Mismatches: 4
Query Match: 99.66% Indels: 1
DB: 9 Gaps: 0
US-09-502-945-2 (1-2885) x US-10-072-094-91 (1-1122)

Qy 2 GAATTCCTCTGTGCAAGTCAAGAGGACCCACACAGCGCGCTCAACCATTCCTCCCA 61
Db 189 GluPheLeuLeuSerLysSerLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 208
Qy 62 CAGCACCCCAAAATCCTGGGAGGCCACCATGCTCTTTTGGACACAGATTCCTCCCTCCCA 121
Db 209 GlnHisProLysCysTrpGlyAlaHisHisAlaSerLeuAspGlnSerSerProGln 228
Qy 122 AGCGGCCCCCTGGAGCGCTCCCTCTACAACTGCCTTTGGCTGGCGCTACGACAGT 181
Db 229 SerGlyProProGlyThrProProSerTyrLysLeuProLeuProGlyProTyrAspSer 248
Qy 182 CGAGCAGACTTCCCTCCCGCAAAACAGCCTCTGAACCCAACTTGAAGTGCCTCAAG 241
Db 249 ArgAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg 268
Qy 242 CTAACACAGAGGTGGTGGAGCGAGAGGAGTCCCTCTCTGCTGCTGCAAGGATGGACT 301
Db 269 LeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArgLysAspGlyThr 288
Qy 302 GTTATTAGCACCCTTAAAGAGAGAGTGTGTGAGATCATCAGGTCCCGGGCTGGGCGTCG 361
Db 289 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyAlaGlyProGlyAlaSer 308
Qy 362 TCCTGTGTGTAAACAGCGACCCGCTCCGCGCCAGCTCTCCCAACAGCTCCCAACAGCACC 421
Db 309 SerValCysAsnSerAlaProGlySerGlyProSerSerProAsnSerSerHisSerThr 328
Qy 422 ATCGCTGAGAAATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCAG 481
Db 329 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGln 348
Qy 482 CACGAGCCCTCCTCTGGACAGCTCCCAACAGCTTACAGCTCTACAGTCTCCTTCT 541
Db 349 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTyrThrSerProSer 368
Qy 542 CTGCCCCAACATCTCCCTAGGCTGAGCGCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 601
Db 369 LeuProAsnIleSerLeuGluGlnAlaThrValThrValThrAsnSerHisLeuThr 388
Qy 602 GCCTCCCGCAAGCTGTGACACAGAGGAGCGCGAGAGCGCGCTCCAGTCCCTCCCGCG 661
Db 389 AlaSerProLysLeuSerThrGlnGlnGluAlaGluArgGlnAlaLeuGlnSerLeuArg 408
Qy 662 CAGGGTGGCAGCTGACCGCAAGTTCATGAGCACATCTCTATTCTGCTGCTGCTGCTGCTGCTG 721
Db 409 GlnGlyGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCysLeuLeu 428
Qy 722 GCGTGGCAGCTGAGGGGACGCGGAGCCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781
Db 429 GlyValAlaLeuGluGlyAspGlySerProHisGlyHisAlaSerLeuLeuGlnHisVal 448
Qy 782 CTGTTGCTGGAGAGCGCGCGCAGCAGCAGCCTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 841
Db 449 LeuLeuLeuGluGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 468
Qy 842 TCCCCACTAGTACGGGTGAACGTGTGGCCACAGCATGCGGACGCTAGGACAGTCCCGCG 901
Db 469 SerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLysLeuPro 488
Qy 902 CGGCATCGGCCCTGAGCGCAGCTCAGTCTCACCCCTGCGCAGAGTCCCCAGGCGCTG 961
Db 489 ArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGlnAlaLeu 508
Qy 962 CAGCAGCTGCTCATGCAACACAGCAGCAGCTCTCTGAGAGAGCAGAGCAGCAGCAGCAG 1021

Db 509 GlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnLysGlnGln 528
QY 1022 CTACAGCTGGGACAGATCCTCACCAAGACAGAGGAGCTGCCAGGACGCCACACCCAC 1081
Db 529 LeuGlnLeuGlyLysLeuThrLysThrGlyGluLeuProArgGlnProThrHis 548
QY 1082 CCTGAGGAGACAGAGGAGCTGACGAGACAGAGGAGGCTTCTGCTGGGGAGGAGGCC 1141
Db 549 ProGluGluThrGluGluGluLeuThrGlnGlnGlnValLeuLeuGlyGluGlyAla 568
QY 1142 CTGACCATGCCCCGGGAGGCTCCACAGAGAGTGAACAGACACAGAGAGCTGGAGAG 1201
Db 569 LeuThrMetProArgGluGlySerThrGluSerGluSerThrGlnGlnLysLeuGlu 588
QY 1202 GAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1261
Db 589 GluAspGluGluGluAspGlyGluGluGluGluAspCysIleGlnValLysAspGlu 608
QY 1262 GCGCAGAGTGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1321
Db 609 GlyGluSerGlyAlaGluGluGlyProAspLeuGluProGlyAlaGlyTyrlsLys 628
QY 1322 CTGTTCTCAGATGCCAACCGTCCAACTTTGACAGGTGTACCAAGCGCCCTCAGCCTG 1381
Db 629 LeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyrlsGlnAlaProLeuSer 648
QY 1382 GCCACTGTGCCCCACAGCCCTGGCGGTACCCCAATCTCCCTGCTGCCCCCTGGGGGC 1441
Db 649 AlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaProGly 668
QY 1442 ATGAGAAGCCCCCAGACCAACCCCTCAAGCACCTTCCACACAAAGTGTGCTACAC 1501
Db 669 MetLysSerProProAspGlnProValLysHisLeuPheThrGlyValValTyrlsAsp 688
QY 1502 ACGTTTCATGCTAAAGCACCAAGTGCATGTGCGGGAACACACAGTGCACCCCTGAGCAT 1561
Db 689 ThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGluHisAla 708
QY 1562 GCGCGATCCAGAGCATCTGCTCCGGCTGCGAGAGACAGGCGCTCTAGCAAGTGCAG 1621
Db 709 GlyArgIleGlnSerIleTrpSerArgLeuGlnGluThrGlyLeuLeuSerLysCysGlu 728
QY 1622 CGGATCCGAGGTGCGAAAGCCACGCTAGATGATGATCCAGACAGTGCATCTGAATAC 1681
Db 729 ArgIleArgGlyArgGlyAlaThrLeuAspGluIleGlnThrValHisSerGluTyrls 748
QY 1682 ACCCTGCTCTATGGACCAAGTCCCTCAACCGGCGAGAAGCTAGACAGCAAGAGTTGCTC 1741
Db 749 ThrLeuLeuTyrlsThrSerProLeuAsnArgGlnLysLeuAspSerLysLysLeuLeu 768
QY 1742 GCTCCCATCAGCCAGAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1801
Db 769 GlyProIleSerGlnLysMetTyrlsAlaValLeuProCysGlyIleGlyValAspSer 788
QY 1802 GACACCGTGTGAATGAGATGCATCTCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1861
Db 789 AspThrValTrpAsnGluMetHisSerSerSerAlaValArgMetAlaValGlyCysLeu 808
QY 1862 CTGGAGCTGGCTTCAAGTGGCTGCAGGAGAGCTCAAGAATGGATTTGCCATCATCCGG 1921
Db 809 LeuGluLeuAlaPheLysValAlaAlaGlyGluLeuLysAsnGlyPheAlaIleIleArg 828
QY 1922 CCCCCAGGACACCGCCGAGGAATCCACAGCATGGGATTTCTGCTTCTCAACTCTGTA 1981
Db 829 ProProGlyHisHisAlaGluSerThrAlaMetGlyPheCysPheAsnSerVal 848
QY 1982 GCATCACCGCAAACTCCTACAGCAGAGTGAAGCTGGGCAAGTCTCATCTGCGAC 2041
Db 849 AlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeuIleValAsp 868
QY 2042 TGGGACATTCACCATGGCAATGGCCACCCAGCAGGCGTCTACAATGACCCCTCTGCTC 2101
Db 869 TrpAspIleHisGlyAsnGlyThrGlnGlnAlaPheTyrlsAsnAspProSerValLeu 888

QY 2102 TACATCTCTCTGATCGCTATGACAAAGGGAACCTCTTTCCAGGCTCTGGGGCTCCTGAA 2161
Db 889 TyrIleSerLeuHisArgTyrlsAsnGlnAsnGlyAsnGlyPheProGlySerGlyAlaProGlu 908
QY 2162 GAGTTGGTGGAGACAGCGCTGGGGTACAATGTGAACGTGGCATGGACAGAGGTGTG 2221
Db 909 GluValGlyGlyGlyProGlyValGlyTyrlsAsnValAsnValAlaTrpThrGlyGlyVal 928
QY 2222 GACCCCCCATTTGGAGACAGCTGGAGTACCTTACAGCCCTTACAGGACAGTGGTGATGCCATT 2281
Db 929 AspProIleGlyAspValGluTyrlsLeuThrAlaPheArgThrValValMetProIle 948
QY 2282 GCCCAGAGTCTCACCTGATGTGGTCTAGTCTCCGCCGGTTTGTATGCTGTTGAAGGA 2341
Db 949 AlaHisGluPheSerProAspValValLeuValSerAlaGlyPheAspAlaValGluGly 968
QY 2342 CATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2401
Db 969 HisLeuSerProLeuGlyGlyTyrlsSerValThrAlaArgCysPheGlyHisLeuThrArg 988
QY 2402 CAGCTGATGACCCCTGGCAGGCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2461
Db 989 GlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGluGlyGlyHisAspLeu 1008
QY 2462 ACCGCCATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2521
Db 1009 ThrAlaIleCysAspAlaSerGluAlaCysValSerAlaLeuLeuSerValGluLeuGln 1028
QY 2522 CCCTTGGATGAGGAGCTTTCAGCAAAAGCCCAACATCAACGAGTGGCCAGCTAGAG 2581
Db 1029 ProLeuAspGluAlaValLeuGlnGlnLysProAsnIleAsnAlaValAlaThrLeuGlu 1048
QY 2582 AAAGTCATCAGATCCAGACCAACTGAGCTGAGCTGCTGAGAGTTCGAGAGTTCGCGCTGCTG 2641
Db 1049 LysValIleGluIleGlnSerLysHisTrpSerCysValGlnLysPheAlaAlaGlyLeu 1068
QY 2642 GCGCGTCCCTGCGAGGCGCCAGCAGGAGTGAACCAAGCAAGCCGAAAT-GTGAACGCC 2700
Db 1069 GlyArgSerLeuArgGluAlaGlnAlaGlyGluThrGluAlaGluThrValSerAla 1088
QY 2701 ATGCGCTTGTGTTGGTGGGCGCAACAGCCCAAGCTGCGCAGCCCGGGAACACAGC 2760
Db 1089 MetAlaLeuLeuSerValGlyAlaGluGlnAlaGlnAlaAlaAlaAlaArgGluHisSer 1108
QY 2761 CCCAGCGCGCAGAGGAGCCCATGGAGCAGGAGCTGCCCTG 2802
Db 1109 ProArgProAlaGluProMetGluGlnGluProAlaLeu 1122

RESULT 3
US-09-817-913-9
; Sequence 9, Application US/09817913
; Patent No. US20020061860A1
; GENERAL INFORMATION:
; APPLICANT: Li, Zuomei
; APPLICANT: Bonfills, Claire
; APPLICANT: Besterman, Jeffrey
; TITLE OF INVENTION: Inhibition of Specific Histone Deacetylase Isoforms
; FILE REFERENCE: 106101.145
; CURRENT APPLICATION NUMBER: US/09/817,913
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/192,157
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Human
US-09-817-913-9
Alignment Scores: 5.5e-170 Length: 717
Pred. No.: 1122

Score: 3612.50 Matches: 706
Percent Similarity: 98.75% Conservative: 3
Best Local Similarity: 98.33% Mismatches: 6
Query Match: 67.57% Indels: 4
DB: 10 Gaps: 3

US-09-502-945-2 (1-2885) x US-09-817-913-9 (1-717)

```
QY 656 CTGGGAGGCTGGACGCTGACGGCAAGTTCATGAGCACATCCTCTATTCTGGCTGC 715
Db 1 LeuArgGlnGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCys 20
QY 716 CTCGTGGGCTGGCACTGGAGGCGCGAGGAGCCGCCACGGCATGCTCCCTGCTGCAG 775
Db 21 LeuLeuGlyValAlaLeuGluGlyAspGlySerProHisGlyHisAlaSerLeuLeuGln 40
QY 776 CATGTGCTGTGTGGAGAGGCGCGGACGACGAGACCCCTCATTTGCTGGCCACTCCAC 835
Db 41 HisValLeuLeuLeuGluGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHis 60
QY 836 GGCAGTCCCTACTAGTACGGGTGAACGTGTGGCCACCATGCGGAGCGGTAGGCAAG 895
Db 61 GlyGlnSerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLys 80
QY 896 CTCGGCGGCTCGGCCCTGAGCCGCACTAGTCTCTACCGTGGCGCAGAGTCCCCAG 955
Db 81 LeuProArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGln 100
QY 956 GCCTGTACAGCTGGGCAAGATCCTCACCAGACAGGGAGCTGCCAGGCGAGCCACC 1015
Db 101 AlaLeuGlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnLysGln 120
QY 1016 CAGCAGCTACAGCTGGGCAAGATCCTCACCAGACAGGGAGCTGCCAGGCGAGCCACC 1075
Db 121 GlnGlnLeuGlnLeuGlyLysIleLeuThrLysThrGlyGluLeuProArgGlnProThr 140
QY 1076 ACCACCTGTAGAGACAGAGGAGGAGCTGACGGAGCAGCAGGAGTCTGTGGGGAG 1135
Db 141 ThrHisProGluGluThrGluGluGluLeuThrGluGlnGlnValLeuLeuGlyGlu 160
QY 1136 GGAGCCTGACCATGCCCGGAGGCTCCACAGAGAGTGAGACACACAGGAACCTTG 1195
Db 161 GlyAlaLeuThrMetProArgGluGlySerThrGluSerGluSerThrGlnGluAspLeu 180
QY 1196 GAGGAGGAGGACGAGGAGGAGTGGGAGGAGGAGGATTCATCCAGGTTAAGGAC 1255
Db 181 GluGluGluAspGluGluAspGly---GluGluGluAspCysIleGlnValLysAsp 199
QY 1256 GAGGAGGCGAGAGTGTGCTGAGGAGGGCCCGACTTGGAGGAGCTGGTGTGGATAC 1315
Db 200 GluGluGlyGluSerGlyAlaGluGluGlyProAspLeuGluGluProGlyAlaGlyTyr 219
QY 1316 AAAAACTGTTTCAGATGCCAACCGCTGCACCTTTCAGGTTACCAAGCGCCCTC 1375
Db 220 LysLysLeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyrGlnAlaProLeu 239
QY 1376 AGCCTGGCCACTGTGCCCCACCAAGCCCTGGGCGGTACCAATCCCTCCCTGCTGCCCT 1435
Db 240 SerLeuAlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaPro 259
QY 1436 GGGGGATGAGAACCCCCCAGACCAACCGCTCAAGCACCTCTTCAACCAAGTGTGGTC 1495
Db 260 GlyGlyMetLysSerProProAspGlnProValLysHisLeuPheThrThrGlyValVal 279
QY 1496 TAGCACACCTTATGCTAAGACACAGTCATGTGGGGGAGAACACACACACCTGAG 1555
Db 280 TyrAspThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGlu 299
QY 1556 CATGCTGGCGGATCCAGAGCATCTGGTCCCGCTGCAGGACAGAGCCTGTAGCAAG 1615
Db 300 HisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnGluThrGlyLeuLeuSerLys 319
QY 1616 TCGAGCGGATCCGAGGTGCGAAAGCCACGCTAGATGATCCAGACAGTGCACCTCTGNA 1675
```

```
Db 320 CysGluArgIleArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGlu 339
QY 1676 TAC---CACACCCTGCTCTATGGGACCAGTCCCTCAACCGGCAGAAAGTAGACAGCAAG 1732
Db 340 TyrIleHisThrLeuLeuTyrGlyThrSerProLeuAsnArgGlnLysLeuAspSerLys 359
QY 1733 AAGTTGCTCGGTCCCATCAGCCAGCAAGATGTATGTGTGCTCCTTGTGGGGGCATCGGG 1792
Db 360 LysLeuLeuGlyProIleSerGlnLysMetTyrAlaValLeuProCysGlyIleGly 379
QY 1793 GTGGCAGTGACACCGCTGTGGAATGAGATGCACCTCCTCCAGTGTGTGGCATGCGACAG 1852
Db 380 ValAspSerAspThrValTrpAsnGluMetHisSerSerSerAlaValAlaArgMetAlaVal 399
QY 1853 GCGTGCCTCTCGAGCTGGCGCTTCAAGTGTGCTGAGGAGCTCAAGAATGATTGCTCC 1912
Db 400 GlyCysLeuLeuGluLeuAlaPheLysValAlaAlaGlyGluLeuLysAsnGlyPheAla 419
QY 1913 ATCATCCGGCCCCCAGGACACCCAGGAGTATCCACAGCATGGGATTTCTGCTTCCTC 1972
Db 420 IleIleArgProProGlyHisHisAlaGluGluSerThrAlaMetGlyPheCysPhe 439
QY 1973 AACTCTGTAGCCATCACCGCAAACTCCTACACAGAAAGTTGAACGTGGCAAGTCTCTC 2032
Db 440 AsnSerValAlaIleThrAlaLysLeuLeuGlnLysLeuAsnValGlyLysValLeu 459
QY 2033 ATCTGGTGGAGTGGACATTCACCATGGCAATGGCACCCAGCAGCGCTTCTACAATGACCC 2092
Db 460 IleValAspTrpAspIleHisHisGlyAsnGlyThrGlnGlnAlaPheTyrAsnAspPro 479
QY 2093 TCTGTGCTCTACATCTCTGTCATCGCTATGACACGGGAACTTCTTCCAGGCTCTGGG 2152
Db 480 SerValLeuTyrIleSerLeuHisArgTyrAspAsnGlyAsnPhePheProGlySerGly 499
QY 2153 GCTCTGTGAAGAGTGTGGAGGACCGGTGGGTACAATGTCAACGTGGCATGGGACA 2212
Db 500 AlaProGluGluValGlyGlyGlyProGlyValGlyTyrAsnValAsnValAlaTrpThr 519
QY 2213 GAGGTGTGGACCCCGCATTTGGAGACGTGGAGTACCTTACAGCCTTCAGACAGTGTGTG 2272
Db 520 GlyGlyValAspProIleGlyAspValGluTyrLeuThrAlaPheArgThrValVal 539
QY 2273 ATGCCATTGCCACGAGTCTCTCCTCATGTGTGTC---CTAGTCTCCCGCGGTTTGTAT 2329
Db 540 MetProIleAlaHisGluPheSerProAspValValThrLeuValSerAlaGlyPheAsp 559
QY 2330 GCTGTGTGAAGACATCTCTCTCTCTGGTGGCTACTCTGTACCCCGCCAGATGTTTGGC 2389
Db 560 AlaValGluGlyHisLeuSerProLeuGlyTyrSerValThrAlaArgCysPheGly 579
QY 2390 CACTTGCACGACGCTGATGACCTGGCAGGGGGCGGGTGTGCTGGCCCTGGAGGGA 2449
Db 580 HisLeuThrArgGlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGluGly 599
QY 2450 GGCCATGACTTGCACCGCCATCTGTGATGCCCTTGAAGCTTGTGTGCGCTCTGCTCAGT 2509
Db 600 GlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValSerAlaLeuLeuSer 619
QY 2510 GTAAAGCTGCAGCCCTTGGATGAGCAGTCTTGCAGCAAAAGCCCAACATCAACGCACTG 2569
Db 620 ValGluLeuGlnProLeuAspGluLeuValLeuGlnGlnLysProAsnIleAsnAlaVal 639
QY 2570 GCCAGCTTAGAAGATCATCGATCCAGACAAACACTGGAGCTGTGTGTGCAAGATTC 2629
Db 640 AlaThrLeuGluLysValIleGluThrGlnSerLysHisTrpSerCysValGlnLysPhe 659
QY 2630 GCCGTGTGTGGCGGTCTCCCTGGAGGGGCCCAAGCAGGTGAGACCGAAGAGCCGAA 2689
Db 660 AlaAlaGlyLeuGlyArgSerLeuArgGluAlaGlnAlaGlyGluThrGluGluAlaGlu 679
QY 2690 AT-GTGAACGCATGGCCCTTCTGTTGGTGGGCGCCGACAGGCCCAAGCTGCGCGACCC 2748
|||||
```

Db 680 ThrValSerAlaMetAlaLeuLeuSerValGlyAlaGluGlnAlaGlnAlaAlaAla 699
Qy 2749 CGGAGACACCCAGCCGCGCAGAGAGCCCATGGACGAGGACCTGCCCTG 2802
Db 700 ArgGluHisSerProArgProAlaGluProMetGluGlnGluProAlaLeu 717
RESULT 4
US-09-817-538-9
: Sequence 9, Application US/09817538
: Patent No. US20020137162A1
: GENERAL INFORMATION:
: APPLICANT: Li, Zuomei
: APPLICANT: Bonfils, Claire
: APPLICANT: Besterman, Jeffrey
: TITLE OF INVENTION: Antisense Oligonucleotide Inhibition of Specific Histone
: TITLE OF INVENTION: Deacetylase Isoforms
: FILE REFERENCE: 106101.144
: CURRENT APPLICATION NUMBER: US/09/817,538
: CURRENT FILING DATE: 2001-03-26
: PRIOR APPLICATION NUMBER: US 60/192,157
: PRIOR FILING DATE: 2000-03-24
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: Patent in version 3.1
: SEQ ID NO 9
: LENGTH: 717
: TYPE: PRT
: ORGANISM: Human
US-09-817-538-9
Alignment Scores:
Pred. No.: 5,5e-170 Length: 717
Score: 3612.50 Matches: 706
Percent Similarity: 98.75% Conservative: 3
Best Local Similarity: 98.33% Mismatches: 6
Query Match: 67.57% Indels: 4
DB: 10 Gaps: 3
US-09-502-945-2 (1-2885) x US-09-817-538-9 (1-717)
Qy 656 CTGCGCAGGGTGGCAGCTGACCGCAGTTCATGACACATCTCTATTCTGCTGC 715
Db 1 LeuArgGlnGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCys 20
Qy 716 CTGCTGGCGTGGCACTGGAGGGCAGCGGAGCCCGCCAGCATGCTCCCTGCTGCAG 775
Db 21 LeuLeuGlyValAlaLeuGluGlyAspGlySerProHisGlyHisAlaSerLeuLeuGln 40
Qy 776 CATGTGCTGTGCTGGAGCAGCCCGGCAGCAGACACCTCATTTGCTGTGCCATCCAC 835
Db 41 HisValLeuLeuGluGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHis 60
Qy 836 GGGCAGTCCCACTAGTGACGGGTGAACGTGTGGCCACCAGCATCGGCAGGTAGCAAG 895
Db 61 GlyGlnSerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLys 80
Qy 896 CTCCCGCGCATCGGCCCTGAGCCGCACTCAGTCTCACCCTGCGCAGAGTCCCGCAG 955
Db 81 LeuProArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGln 100
Qy 956 GCCTCGACAGCTGTGTATGCAACACAGCAGCAGCAGCATTTCTTGGAGACGAGCAG 1015
Db 101 AlaLeuGlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnLysGln 120
Qy 1016 CAGCAGTACAGTGGCAGAGTCTCCACCAAGCAGAGGGAGCTGCCAGCAGCCACCC 1075
Db 121 GlnGlnLeuGlnLeuGlyLysIleLeuThrLysThrGlyLeuProArgGlnProThr 140
Qy 1076 ACCCACTCTGAGGAGACAGAGGAGTGTGAGGAGCAGCAGAGTCTTGTGGGGAG 1135
Db 141 ThrHisProGluThrGluGluGluLeuThrGluGlnGlnGluValLeuLeuGlyGlu 160
Qy 1136 GGAGCCCTGACCATGCCCGGGAGGGCTCCACAGAGAGTGTGAGCAGCAGCAGTGTG 1195

Db 161 GlyAlaLeuThrMetProArgGluGlySerThrGluSerGluSerThrGlnGluAspLeu 180
Qy 1196 GAGGAGGAGCAGCAGAGAGGATGGGAGGAGGAGGATTCATTCAGGTTAAGGAC 1255
Db 181 GluGluGluAspGluGluAspGly---GluGluGluAspCysIleGlnValLysAsp 199
Qy 1256 GAGGAGGCGAGAGTGTGCTGAGGAGGGCCGACTTGGAGGAGCTGGTGTGGATAC 1315
Db 200 GluGluGlySerGlyAlaGluGluGlyProAspLeuGluGluProGlyAlaGlyTyr 219
Qy 1316 AAAAATGTTCTCAGATGCCCAACCGCTGCAACCTTTGAGGTGTACCAAGGCCCTC 1375
Db 220 LysLysLeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyrGlnAlaProLeu 239
Qy 1376 AGCTGCGCACTGTGCCCAACCAAGCCCTGGCGCTGACCAATCTCCCTGCTGCCCT 1435
Db 240 SerLeuAlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaPro 259
Qy 1436 GGGGCGATGAAGAACCCCGCCAGACCAACCGTCAAGCACCTCTTCCACCAAGTGTGGTC 1495
Db 260 GlyGlyMetLysSerProProAspGlnProValLysHisLeuPheThrThrGlyValVal 279
Qy 1496 TAGACACGTTTCATGTAAAGCACCAGTGCATGTGGGGGAACACACACGTCACCTGAG 1555
Db 280 TyrAspThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGlu 299
Qy 1556 CATGCTGGCGGATCCAGACATCTGGTCCCGCTGCAGGAGACAGCCCTGTAGCAAG 1615
Db 300 HisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnGluThrGlyLeuLeuSerLys 319
Qy 1616 TGGGAGCGGATCCGAGTGCAGAACCCACGCTAGATGAGATCCAGACAGTGCCTGAA 1675
Db 320 CysGluArgIleArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGlu 339
Qy 1676 TAC---CACACCTGCTATGGACCCAGTCCCTCAACCGCAGAGAGTAGCAGCAAG 1732
Db 340 TyrIleHisThrLeuLeuTyrGlyThrSerProLeuAsnArgGlnLysLeuAspSerLys 359
Qy 1733 AAGTGTGCTGCTCCATCAGCCAGAGATGTATGCTGTGCTGTGCTGGGGCATCGGG 1792
Db 360 LysLeuLeuGlyProIleSerGlnLysMetTyrAlaValLeuProCysGlyGlyIleGly 379
Qy 1793 GTGACAGTGCACCCGTGTGGATGAGATGAGATGCTCCTCCAGTGTGCTGCGCAGTGCAGT 1852
Db 380 ValAspSerAspThrValTrpAsnGluMetHisSerSerAlaValArgMetAlaVal 399
Qy 1853 GGCTGCTGTGGAGCTGGCTTCAAGTGGCTGCAGAGAGCTCAAGATGGATTTGCC 1912
Db 400 GlyCysLeuLeuGluLeuAlaPheLysValAlaAlaGlyGluLeuLysAsnGlyPheAla 419
Qy 1913 ATCATCCGGCCCCAGGACACCCAGCCGAGGAATCCACAGCATGGGATTTCTTCTTC 1972
Db 420 IleIleArgProProGlyHisHisAlaGluGluSerThrAlaMetGlyPheCysPhePhe 439
Qy 1973 AACTCTAGCCATCACCGCAAACTCCTACAGCAGAGTGAACGTGGGCAAGTGCCTC 2032
Db 440 AsnSerValAlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeu 459
Qy 2033 ATCTGTGACTGGGACATTCACCATGGCAATGGCACCCAGGCGGTCTACAATGACCCC 2092
Db 460 IleValAspTrpAspIleHisHisGlyAsnGlyThrGlnGlnAlaPheTyrAsnAspPro 479
Qy 2093 TCTGTGCTCTACATCTCTCTCATCGTATGACAAACGGGAACCTTTTCCAGGCTCTGGG 2152
Db 480 SerValLeuTyrIleSerLeuHisArgTyrAspAsnGlyAsnPhePheProGlySerGly 499
Qy 2153 GCTCCTGAAGAGTGTGGAGGACCGGCTGGGTGTACATGTGAACCTGGCATGGACA 2212
Db 500 AlaProGluGluValGlyGlyProGlyValGlyTyrAsnValAsnValAlaTrpThr 519
Qy 2213 GGAGGTGTGGACCCCTTGGAGACGTGGAGTACCTTACAGCTTCAGCAGTGTGGTGTG 2272
Db 520 GlyGlyValAspProIleGlyAspValGluTyrLeuThrAlaPheArgThrValVal 539